THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME VII (1940)

GENERAL EDITOR
J. H. AWBERY

Price to non-Fellows 22s. 6d. net

Reproduced and Printed Photo-Litho Offset for
WM. DAWSON & SONS LTD.
16, West Street, Farnham, Surrey
1960

With the permission of
THE PHYSICAL SOCIETY

CONTENTS

PAGE

Sound. By E. G. Richardson 1

Musical acoustics. By E. G. Richardson 27

Photo-electric spectrophotometry and its use in industry. By W. D. Wright . 36

The infra-red spectra of polyatomic molecules. By Harald H. Nielsen . . 41

The gamma radiations emitted in nuclear processes. By N. Feather . . 66

The Bohr theory of nuclear reactions. By R. Peierls 87

Electron microscopes of high magnification. By O. Klemperer . . . 107

New lens systems. By H. W. Lee 130

The absorption of light in interstellar space. By A. Hunter 150

Solar physics. By A. D. Thackeray 160

Surface tension. By R. C. Brown 180

Equations of state. By James A. Beattie and Walter H. Stockmayer . . 195

Electrical discharges in gases and their applications: Part I. By V. J. Francis and H. G. Jenkins 230

Some interactions of gases with metals and crystalline solids. By J. K. Roberts . 303

Viscosity and the nature of substances of high molecular weight in solution. By F. Eirich 329

The development of the teaching of experimental physics in British universities. By Allan Ferguson 355

SOUND

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.,
King's College, Newcastle-upon-Tyne

§1. FREE PROPAGATION: SUPERSONICS

APPARATUS for the measurement of velocity and attenuation of sound and supersonics is now fairly established, and in default of the discovery of an entirely new technique we must expect a diminution in the number of papers which deal with methodology. Hubbard and Zartman (1939) have introduced a modification of the acoustic interferometer for gases in which the distance between source and reflector remains fixed (apart from the effects of thermal expansion). This type of instrument is particularly of use when it is difficult or impossible to provide the reflector with a micrometer screw actuated from outside, e.g., in studying the propagation of supersonics in gases at low temperatures or high pressures. One calculates the propagation constants in the gaseous column by considering the impedance of the system—quartz plus gas-path plus reflector. If the values of the pressure and temperature are so chosen that a resonant standing-wave system is set up, the calculation is facilitated as the authors show: the required constants result from observations of the size and disposition of the reaction peaks in the current in the maintaining circuit. Alleman (1939) deals further with the point raised by Herzfeld (cf. last *Report*) by which results for attenuation constants derived from such impedance methods may be in error due to a neglect of the heat losses by conduction at the reflector surface. He gives a mathematical analysis of such an interferometer with the inclusion of terms representing the effects of this new form of "dissipation reflection" too involved to be reproduced here. He discusses not only how a correction factor may be applied to the results of Hubbard and others who have used the impedance method, but also the desiderata to be looked for in a good interferometer of this type.

Passing now to actual results for gases; Kneser (1939) and also Pielemeier (1939) have been concerned about the "spread" of the results of various workers who have experimented with dry air and reduced their results to 0° C. While the former thinks the discrepancy is due to friction of the sound waves along the walls of the containing vessel lowering the velocity more or less below the "free" value, the latter groups the most accurate values which have been obtained under various conditions, indoor and outdoor, sonic and supersonic, and shows that even these carefully selected data range from 330·6 to 331·9 m./sec. after reduction. He considers that the discrepancy is largely to be attributed to the presence of

water vapour and carbon dioxide in the atmosphere or in the experimental chamber in various proportions. The investigators in most instances reduce their results by making the reduction one of density merely, whereas they ought to allow for the vapour pressures of the two adulterants mentioned, small quantities of which may cause some dispersion of the velocity. Schulze (1939) carries out a painstaking research on chlorine at audible pitch, using resonance tubes a foot or so in length and about an inch wide, mounted in a thermostat and driven by a telephone diaphragm. The attainment of resonance conditions was assured by ear. The Helmholtz-Kirchhoff correction for tube-width was applied. Measurements of velocity at a series of values of temperature showed dispersion, the wave-lengths at the high frequencies being super-normal. Plotted against the parameter; frequency/pressure, the velocity remains constant up to the point where, under atmospheric pressure, the frequency would be 12,000 kc./sec. and then rises by about 2 per cent as the frequency transcends the upper audible limit. Values deduced for the dependence of specific heat on temperature in chlorine agree with those of Eucken and Becker.

Van Itterbeek, de Bruyn and Mariens (1939) have made further measurements on carefully purified carbon dioxide together with some on the same gas with small adulterations of oxygen, hydrogen, deuterium and the two kinds of water vapour at 600 kc./sec., though their apparatus still involves the pressure on an obstacle of considerable size, an undesirable feature when absorption coefficients in a gas are being determined. It is interesting that, though they find, with other workers in this field, attenuation greater than that to be expected on classical theory, the addition of oxygen to the carbon dioxide lowers the experimental value eventually to concordance with the theoretical one.

There are two investigations on organic vapours to report, both of which employ a Pierce interferometer in a sealed chamber initially evacuated and then filled with the vapour at an appropriate pressure. Out of eleven vapours tested by Railston (1939) at frequencies from 25 to 1000 kc./sec. only two, carbon disulphide and benzene, showed a dispersion exceeding two per cent. (In the two exceptions, the variation in velocity ranged over six and four per cent respectively.) Benzene, too, showed a fall below the audio-frequency value between 25 and 100 kc./sec. with a subsequent recovery. Abnormal absorption coefficients were also noted in those substances which show excessive absorption in the liquid state. Jatkar's measurements relate to velocity only (Jatkar, 1939). He has measured supersonic wave-lengths over a more restricted range—50 to 125 kc./sec.—not only in air, steam and carbon dioxide, but also in about a score of organic vapours including also benzene and carbon disulphide (the latter at 50 kc./sec. only). No dispersion was found in this region. Jatkar's results cover, however, a range of temperature up to 135° C.

Coming now to experiments which touch on the theory of supersonic dispersion, Dwyer (1939) describes some in which the after-effect of collisions produced when a discharge passes through iodine is evidenced by the persistence

of absorption lines in the visible spectrum for a short time after the discharge has ceased. Such after-effects were postulated in the relaxation-time theory in its original form, though their existence in Dwyer's demonstration has no necessary connection with supersonic dispersion since Dwyer made no experiments with such radiation. The writer (Richardson, 1939) has tried to find a direct effect on the well-known supersonic dispersion in carbon dioxide of subjecting it to optical radiation of its own natural (vibrational) frequencies, which are located in the infra-red region of the spectrum. Supersonic frequencies between 90 and 100 kc./sec. were used with the gas in a Pierce interferometer. The infra-red radiation came from a hot tube filled with carbon dioxide and was focussed on to one of two quartz windows in the interferometer tube. The beam emerging from the opposite window was tested in an infra-red spectrometer to see whether the supersonic rays had affected it, but without positive result. On the other hand, the supersonic beam itself was found to be more attenuated when the gas was irradiated by the infra-red than when it was not, as though the molecules of the gas, excited in their natural vibrational states, were able to absorb more supersonic energy.

Continuing their efforts to improve the technique for measuring supersonic wave-lengths in liquids, Fox and Rock (1939) have devised a simple modification of the Bär system of light diffraction which appears to be simple to set up and, according to the authors, is capable of high precision. No lenses or slits are used, the light diverges from a line source, passes athwart a sound beam which modulates the light waves, then across a second sound beam of like frequency traversing the liquid under observation in its trough. The progressive waves in the liquid can then be observed stroboscopically by means of the modulated light beam, the optical system being precisely the same as that used in the familiar method of measuring the wave-lengths of ripples on the surface of water. The bands of compression may be observed directly or cast as shadows on a photographic plate. Standing waves in the trough are eliminated by a "complete absorber" consisting of the diffusing chamber illustrated in the last *Report*. A feature of the optical system is its simplicity, which makes adjustment for plane-parallel light unnecessary and allows of exact calculation of the wave-length in terms of geometrical constants which can be determined with great accuracy.

An optical method has been used by Liepman (1939) for wave-lengths in liquid argon and nitrogen. The diffraction spectrum is produced when light is reflected to and fro across a cell of the liquid (contained in a Dewar flask) irradiated by a supersonic source. In this way the velocity as function of the temperature is derived, both for the two liquids severally and for mixtures in various proportions. Both he and Groenewald (1939)—in relation to helium II—consider the possibility of a failure of adiabatic propagation, but reach the conclusion that this cannot happen within the range of attainable supersonic frequencies. The possibility of association of the molecules as an alternative explanation of the results is not rejected out of hand.

Zachoval (1939), Bhagavantam and Rao (1939), Krishnan (1939) and Salceanu (1939) have used organic liquids. The first-named claims a dispersion of one per cent between the first and third harmonics of a source of 2000 kc./sec. working in resin oil, while toluene, nitro-benzene and glycerine show a fall of velocity between 1000 and 3000 kc./sec., though only of the order of one or two per cent. The Indian investigators, too, use higher frequencies (from 1 to 13 Mc./sec.) in benzene, carbon tetrachloride and disulphide and glycerine. Their results show no certain dispersion in this region in either carbon tetrachloride or distilled water. In the latter Krishnan attained the colossal frequency of 22·5 Mc./sec. without noticing a change in velocity. But a number of other writers consider that the small amount of dispersion which they have noted in liquids is rather to be ascribed to technical difficulties in carrying out the experiment. Both Matossi (1939) and Schreuer (1939) remark that the velocity—and so too the dispersion as between values at different frequencies—is a function of intensity. The latter can be raised through the coupling between the quartz source and the valve oscillator which maintains it. When this was done (in water or xylol) with two sources of 3000 and 10,000 kc./sec. respectively, "dispersions" of velocity with frequency could be produced greater than any observable between the results for the pair of sources under equal coupling. Using the Hiedemann technique, Schreuer also found that velocity depends on intensity of source. Other troubles which affect the apparent velocities were traced to distortions in the sound field caused by irregular temperature effects and resulting departure of the wave-fronts from plane form. On extrapolating his results to zero intensity, he could establish no dispersion greater than 0·01 per cent. Curiously enough, the velocity in water rises by 0·04 per cent as the intensity of the source is increased by a factor of five, but in several organic liquids it goes down as the power supplied to the source is raised.

Field (1939 a) reverts to an earlier experiment in which he showed that longitudinal waves sent through a column of liquid in a thin-walled tube might be distorted by radial resonance. Repeating the experiment in a thick-walled tube, he finds no such distortion. Evidently, unless the walls give under the radiation and so encourage radial oscillations of the liquid column, longitudinal waves sent in at one end will remain plane; on the other hand, if these radial vibrations do intervene, the system acts as a filter for these resonant frequencies.

Biquard (1939) continues his work on the absorption of supersonics in liquids by examining the effect of pressure on this factor. He likewise practises the method of light diffraction and is able to conduct the work in a chamber in which the liquid may be submitted to a pressure of 1000 kg_i per sq. cm. The vessel is a hollow steel cylinder nearly a metre long and 65 mm. internal diameter, provided with a number of pairs of facing windows throughout its length, through which the light beam is directed. The supersonic amplitude at the various stations is measured by the intensity of the light diffracted. For this a photo-electric cell is used, but it is not shown why the light intensity at each window should bear a linear relation

to the supersonic amplitude in question. The end of the tube remote from the quartz emitter has jambs set at an acute angle to the axis of the tube so as to throw the reflected beam against the walls of the tube and not confound the progressive wave system in which the amplitudes are being measured. Results are given for benzene, toluene, carbon tetrachloride, methyl acetate and water, each at one frequency and for pressures mounting up to 630 kg. per sq. cm. In the liquids studied, the velocity rises nearly in linear fashion with the pressure, while the coefficient of absorption runs in the opposite sense, in accordance with the known change of viscosity with pressure (Bridgman's results).

Rieckman (1939) measures supersonic absorption in mercury using the torsion vane method. Even so, it was necessary to suspend the vane in another liquid (water) and to let the energy which had traversed a given stretch of mercury fall upon a window and thence into the water to fall upon the torsion arm. The material of the window must have a specific impedance intermediate between that of mercury and that of water. Actually mica was the substance chosen as fulfilling this requirement. Nevertheless, the window had a reflection coefficient of about one-half, and the energy sent back had to be absorbed. This was done by putting a coating of pertinax paper over the surface of the iron trough containing the mercury. Around the walls of the chamber holding the torsion balance, a rubber sheet served a similar purpose. The divergence of the beam from the forthright direction from the quartz is investigated by the author. Of course, the absorbent lining is helpful here too, but it is as well to remember that, on account of the greater ratio of wave-length to aperture of source, such divergence with consequent missing of part of the beam by the detector—in this case, the vane—is a more frequent cause of error in absorption measurements in liquids than in gases. The frequencies used by Rieckman consisted of the series of harmonics of the fundamental 7·5 Mc./sec. lying between 20 and 55 Mc./sec.; the ratio of absorption coefficient (α) to the square of the frequency (n) remained constant at $1{\cdot}3 \times 10^{-16}$.

Sata and Namu (1939) continue to study the influence of supersonics upon certain properties of colloidal solutions, such as the thixotropy of aluminium hydroxide and the optical rotation afforded by gelatine solution. Freundlich and Gillings (1939) observe—what many people must have already surmised—that in many of the "effects" of this class so quickly discovered when this new tool became available a decade ago, there is nothing specific in the action of the very high frequency. Given equal dispensations of energy many of the new phenomena can as well be produced by audible sound waves. Schmid and his collaborators show that the molecular weight of large molecules, such as that of polystyrol, may be lowered by supersonic irradiation (Schmid, Rommel and Roll, 1939). The needle-like crystals of the alloy Wood's metal can also be shortened under this influence during the process of cooling. In each case the action appears to be a frictional one, the molecules or crystals being abraded into smaller fragments by the intense agitation.

There are two supersonic researches of a biological nature to note. Pohlman (1939) places pieces of human tissue in a physiological salt solution between quartz and vane detector in order to measure the absorption afforded by the tissue. α/n^2 is found to diminish as the frequency goes up, though we are not clear as to the precise physiological significance of this result. Hopwood and his co-workers have shown that a vaccinia virus is not reduced in infectivity as a result of exposure to the radiation (Hopwood, Salaman and McFarlane, 1939). Indeed calf-lymph virus has even shown increased activity as a result, presumably, of disengagement under the radiation of the virus proper from inert material found in the crude lymph which otherwise protected the activity of the virus itself.

There are two papers dealing with shock waves. Thompson and Riffolt (1939) discuss the velocity of propagation of such waves and suggest a modification of the Riemann formula; $V^2=c^2(1+k/x^2)$ (c = normal velocity of sound; V = velocity of shock wave at a distance x from the centre of the disturbance), by the introduction of an additional constant ϕ in the form: $V^2=c^2\{1+\kappa/(\phi+x)^2\}$ (x is to be measured in half-diameters of the equivalent spherical source). They have checked this new formula with measurements, employing piezo-electric pressure gauges, of the times of passage of a wave of detonation past a series of stations, and found good agreement with the experiments. For a spherical source ϕ is put equal to 3; 2·75 for a cylindrical source. A discussion ensues of the appropriate value for the ratio of specific heats for condensation cycles of short duration by means of the same theory of relaxation for the transfer of energy between types of molecular vibrations as is familiar in the theory of supersonic propagation in gases. For wave velocities equal to $5c$ or more, it is reckoned that the total heat for a diatomic gas in translation would be 7 and the ratio accordingly 7/5; the same figure would apply to the region of the wave front for velocities only slightly in excess of c. It is assumed that the ratio becomes 5/3 as soon as the molecular trajectories become in essence uni-directional—that is for wave velocities much in excess of $1{\cdot}5c$—but at velocities in excess of $3c$ the vibrational energy ceases to be entirely negligible. So the value of γ decreases slowly again from 5/3 towards 7/5 as the wave surface increases in sharpness and amplitude. This prompts the suggestion of a mechanism by which vibrational energy can build up relatively slowly with a pulse of moderately high velocity. One such idea is that the centre of the zone of reaction lags behind the surface of the shock which it supports, that is, behind the detonating pulse, with the rate of reaction tapering off rapidly from this point towards the boundary of the pulse.

Becker (1939) attempts to analyse sound by making the source set up capillary waves on the surface of mercury and examining the light reflected from this surface in a sound film or photo-electric cell. The ripples have very small amplitudes of the order of 0·1 micron and are excited by a loud speaker, to the armature of which is fixed a knife-edge whose lower end dips into the liquid. It was found that the reflected light intensity rose at first with current through the loud-speaker coil, but later reached a constant value (for frequencies exceeding

1000 c./sec.). Lower frequencies showed a maximum at certain values of the power output to the dipper. In the actual method of analysis, the sound in question is applied to the electrical transducer and so transformed into capillary waves from which the reflected light is caught by a moving-film camera. Specimen records appear as black lines with symmetrical markings on either side where the erstwhile steady reflected beam has been modulated by the ripples. The examples shown include speech, pianoforte and orchestral music, though the actual Fourier analysis of these films does not yet seem to have been carried out.

§ 2. VIBRATING SYSTEMS: SOLIDS

Field (1939 b) derives equations for the resonance radial frequencies of a hollow cylinder having a thick wall in which he takes into account waves reflected to and fro within the confines of the wall, set up by an alternating pressure difference between the inside and the outside of the wall. In an earlier paper, dealing with the thin-walled cylinder, a simple extensional vibration of the tube wall was considered, but this concept will not do when the wall is thick. Results derived from the more elaborate theory now developed are of importance for the natural frequencies of vibration of a quartz resonator cut in the form of a ring.

Shankland and Coltman (1939) give a theory of the vibrating string (in transverse vibration) which follows the treatment of Rayleigh in terms of normal functions. This theory allows of the consideration of a finite amplitude of motion causing a variation in tension which is a function of time, distance along the string and nature of support at the bridges. This variation is treated as a perturbation which modifies the ideal motion of a string with infinitesimal amplitude. The motion of the string under its variable tension being given by the equation

$$\frac{\partial}{\partial x}\left(T\frac{\partial y}{\partial x}\right)=m\frac{\partial^2 y}{\partial t^2},$$

a substitution of the type $y=u_n(x)\exp.(2\pi c f_n t)$ gives a second equation,

$$\frac{\partial}{\partial x}\left(T\frac{\partial u_n}{\partial x}\right)+4\pi^2\rho f_n{}^2 u_n=0.$$

Here f_n is the frequency of vibration for finite amplitude (A), and $u_n(x)$ are functions delineating the shape of the string in each case. The authors performed experiments on the change of frequency with amplitude in a brass wire and found that the results satisfied an equation $f_1{}^2=f_0{}^2-13{\cdot}5A_1{}^2$ which is of the same form as the solution of the last-quoted theoretical equation. When the amplitude is small, stiffness becomes a predominating factor. When this has to be reckoned with they write

$$\frac{\partial}{\partial x}\left(T\frac{\partial u_n}{\partial x}\right)-\frac{\pi r^4 Y}{4}\frac{\partial^4 u_n}{\partial x^4}+4\pi^2\rho f_n{}^2 u_n{}^2=0$$

(r = radius of wire; Y = Young's modulus), and find a solution

$$f_n=\frac{n}{2l}\left(\frac{T}{m}\right)^{\frac{1}{2}}\left[1-\frac{\pi^2 r^2 Y n^2 C}{8l^2 T}A_n{}^2\ldots\right].$$

Dependence of frequency on amplitude under these conditions (of small amplitude) in their experiments exhibits the working of the stiffness factor.

M. Ghosh (1939) continues his applications of operational methods to the theory of the struck string. In the present paper the hammer is allowed to impinge at any point on a string of finite length. Previous analyses on the theoretical side had ignored the coupling between the vibrations of the two segments into which the string is divided at the point of impact; in other words, the effect of successive reflections of the waves from either end had been neglected and only the once-reflected wave considered. In the {displacement, time} curve for any point on the string there are superposed on the large hump, caused by the initial blow, kinks due to disturbances that overtake the hammer from the remote end of the string. These are satisfactorily accounted for in Ghosh's theory. Davy, Littlewood and McCay (1939) have been able to measure the actual force between hammer and string from instant to instant during the short time of contact. A knowledge of the actual instantaneous acceleration of the hammer is of importance for the theory of the struck string, though it is usual among theorists to make an *ad hoc* assumption with regard to this aspect of the problem and to see how it checks with experimental observations on the subsequent motion of the string. A piezo-electric quartz pressure-gauge was placed in the hammer head and the instantaneous force upon it observed in a cathode-ray oscillograph with linear time-base. The areas under the oscillograms are proportional to the corresponding changes in momentum of the hammer so that their ordinates can be taken as proportional to the instantaneous values of the force of contact. The work falls into line with that of Ghosh just described, for the resulting {force, time} curve takes the shape of a hump with a number of overtones or kinks superposed as a ripple upon it.

Passing on to systems exhibiting elastic vibrations, Thomson (1939) contributes a theoretical paper on the transverse vibrations of prismatic bars in which the effects of both rotatory and lateral inertia are included; Colwell (1939) and his collaborators continue their series of papers on the Chladni patterns of circular plates, while Mary Waller (1939) now does the same for the free square plates.

The photographs of the figures for the circular plates show that two normal modes of vibration may subsist at the same time, and this conjunction gives rise to symmetrical figures which are neither diameters nor circles merely. In the case of the square plate any normal mode may be identified by a set of simple rules involving a count of the number of lines across a diameter and along an edge. The amplitude corresponding to any mode is usually sufficient to move the sand over the greater part of the surface. By contrast, on the circular plate when there are many nodal diameters, the sand is only set in considerable motion near the free edge. Recent photographs of the figures by Miss Waller are compared with the corresponding original Chladni drawings.

Jan Arts (1939) continues his intriguing "Jottings from my Experiences with the Sound of Bells", by calling attention to the "secondary strike note". This

is a tone which he localises at one-fourth above the fifth partial tone of the bell and which, starting very loudly, performs a rapid *diminuendo*. Eight seconds is the maximum time of waning that he has noted. This tone appears to correspond to a mode of vibration which has a nodal circle about halfway between the rim and the shoulder of the bell.

§3. VIBRATING SYSTEMS: FLUIDS

It is Meyer and Tamm (1939) who this year provide the paper on a subject which seems to have aroused much interest among research workers in sound in the past few years, i.e., the natural frequency and damping of gas-bubbles formed in liquids. Their bubbles were produced by electrolysis of the liquid in a vessel which also contained a magneto-strictive source of supersonic vibrations (15 to 35 kc./sec.), and it was noted that the frequency of this oscillation determined the size of the bubbles liberated, in such a way that the product of this frequency and the mean diameter of the bubbles was a constant. This is what one would expect if the vibrations of the bubble were radial. It was also possible to see under the microscope when resonant vibrations of the bubble took place, because it then took on a matt instead of a sharply defined appearance. To obtain the actual resonance curves with a view to calculating the damping coefficient the bubble was allowed to grow in size and then, while still irradiated with the source at the original resonance frequency, the reaction due to the oscillation of the bubble upon a nearby microphone in the liquid observed. This gave them a {response, natural frequency} curve for the bubble, on the assumption that the natural frequency of the growing bubble was inversely as its radius at any instant. So the damping coefficient in the neighbourhood of a resonant peak was calculated following the usual theory of forced vibrations. Larger bubbles—in resonance to the range of frequency 1·5 to 7 kc./sec.—had their amplitude of forced vibration directly measured by casting an image on a photo-electric cell, in which an A.C. proportional to the amplitude in question was set up.

The effect of the viscosity of the liquid on the bubble vibrations also came under review. This naturally affected the damping coefficient; the latter also increased in linear fashion with the natural frequency of the bubble. Finally, the sound spectrum in the frequency range up to 10 kc./sec. for the hiss formed by bubbles escaping from a fine jet submerged in water was obtained. Specimens of these are shown in figure 1. The maximum noise level comes from those bubbles (as judged by their natural frequency) which have a diameter equal to that of the orifice. Round this value the {size, frequency} diagram seems to be grouped approximately in the form of a Gaussian error curve. Figure 1 shows three such spectra with jets 1·2, 1·65 and 1·95 mm. diameter respectively. The product of frequency showing maximum intensity and diameter is 660 in each case.

Jarmila Cerovska (1939) has studied the elastic vibrations of the fluid filling the cavity in a quartz ring cut perpendicular to the optic axis, so that the electric

axis lies in the plane of the ring, and set in forced vibration. She used lycopodium powder to exhibit the modes of vibration of the air cavity and also when the space was filled with carbon tetrachloride or xylol. The schlieren and light-diffraction methods were also tried, but did not give such well-defined photographs as the powder. Heating of the liquid was a frequent source of trouble, distorting the symmetry of the figures.

Carrière (1939 a) continues with his favourite subject, the singing flame. He is now considering its possible use as a fog signal. He has not, indeed, made any measurements of the distance to which its sound will carry, but only of the relative pressure-amplitude in a heat-maintained column of air to that in an organ pipe of equal calibre. He finds that one can get an output of sound energy up to five

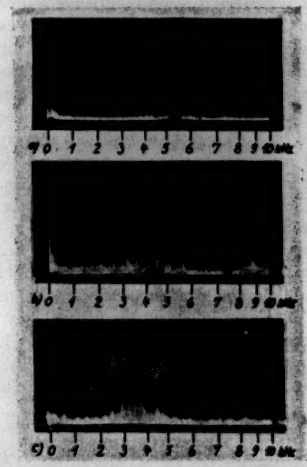

Figure 1. Sound spectra of submerged air jet (Meyer and Tamm in *Akust. Z.*).

times greater when the former is used, though this naturally depends on the wind pressure upon which the organ pipe stands. It appears from his description that the source used by Carrière was not the conventional singing flame, but one of the Lissajous type in which a bunsen burner is placed near the lower open end of the column, in distinction to that in which the flame is brought by a supply tube to the central node. Pressure amplitudes were measured by a manometric capsule fixed to the side of the tube near its centre. The author notes a regular intermission of the sound which, if it could be controlled, would be useful in fog signalling. This interruption appears to arise whenever the aerial amplitude rises

to such a value as to drive the gas back into the burner tube, until the resulting diminution of amplitude allows the flame to return to its former position and the sound recommences.

Taber Jones (1939), besides giving a résumé of recent work on organ pipes which has already been described in these *Reports*, deals in a second paper with the natural frequencies of tubes into which a coaxal rod has been partially inserted, making it into a pipe compounded of a cylindrical and an annular tube of the same external diameter. Experiments show that the same formula as for two cylinders of different diameters joined end-to-end applies to this combination, i.e. it is the same as for the organ stop called *flute-à-cheminée* or *röhrflöte*. He treats of this combination too, but with the modification of a connecting cone joining the two cylinders in place of the more usual sudden change of cross-section. Both Auger (1939) and Mokhtar (1939) have done work on reed pipes in which the vibrations of the column of air are compared to those of the reed. Whereas the former is more concerned with the jump in frequency which occurs when the blowing pressure reaches a critical value, the latter has obtained oscillograms of the reed and the pressure variation in the pipe just above. Actually, as Mokhtar's wave-forms show, the quality of the pipe grows gradually in richness of overtone structure as the wind pressure is increased; all the time, the vibrations of the reed itself have remained simple harmonic. When the critical pressure and consequent jump occur, the fundamental disappears from the oscillogram. Mokhtar also discusses the nature of the coupling between reed and air column which he exemplifies on a series of wave patterns which indicate the phase relationship between their respective vibrations, as the column is lengthened (meanwhile retaining the wind pressure constant at a value slightly in excess of the minimum to make the pipe sound). This appears to be of the relaxation type in which the forcing impulse from the reed occurs once in every several vibrations of the air column, the relaxing being evident in the falling amplitude of the subsidiary peaks in the wave-form occurring between the successive openings of the space between the reed and the table against which it beats. As the length of the tube resonator is increased and its natural frequency becomes more and more out of tune with that of the reed, the peak in the air compression lags more and more behind the reed's motion until they are so at sixes and sevens that the system is reduced to silence. By raising the blowing pressure at this stage it may be made to sound again. The fact that with the shorter length of resonator we have a simple harmonic motion on the part of the driver maintaining relaxation oscillations of the air column shows that the motion is maintained by periodic impulses and not by continuous forcing. With periodic impulses of duration short compared to the full period, maintenance is most efficaceous if the impulse of pressure occurs at the epoch when the pressure in the air column is a maximum (cf. the analogous case of the singing flame). The optimum length of tube is that for which the compression just travels to the open end and back in the time between two impulses. As the tube is lengthened the system will pass through a series of favourable

(corresponding to the condition just named) and unfavourable lengths, when the system will generally cease to sound.

Phelps (1939) uses a column of air as a sound field for calibrating a microphone. It consists of an eight-foot iron pipe ten inches in diameter with a dynamic loud-speaker at one end and an adjustable pistoh at the other. The microphone can be carried to any position in the standing wave system so set up on a rod which passes excentrically through the piston. He works out the impedance of the system, and from this the pressure and velocity amplitudes along the pipe. If dissipation be negligible it is always possible to find positions where either the pressure or the velocity amplitude is zero; this will not be exactly so, owing to finite damping in the air column and the fact that the stopper may not reflect all frequencies equally well, but at any rate these amplitudes will be small at one-quarter and one-half wave-length from the end respectively. Actually it was found, by traversing the microphone along the tube, that a node or antinode of a given harmonic was truly a node or antinode in respect of this harmonic within 2 or 3 per cent. At 100 c./sec. a pressure of 1100 dynes/sq. cm. is obtained at the nodes with the speaker working at less than one half-watt input.

Dunn (1939) employs a vacuum thermocouple connected to a Grassot fluxmeter as an instrument for measuring the average power in fluctuating sounds such as noise and speech sounds. The electric currents to which these give rise in a suitable acoustic transducer are passed through the heater of the thermo-element. Since the voltage from the couple is not proportional to the square of the heater current at every instant, it is not at once apparent that the deflection of the fluxmeter measures the total energy in the heater. But over a fairly long time double integration takes place, the heater power being integrated into a fluctuating temperature and this in turn into a fluxmeter deflection. When measuring sound energy, the current is passed through the heater for the specified interval, after which the circuit is opened. Time must then be allowed for the thermocouple to cool and the motion of the fluxmeter coil to be damped. When the needle finally comes to rest the author shows, on the basis of certain theoretical assumptions, that the deflection is proportional to the current supplied to the heater and therefore to the acoustic energy transformed in that time. Tests on actual sounds are not yet reported.

There are still a number of papers annually on that indispensable sound-measuring instrument, the Rayleigh disc. Merrington and Oatley (1939) have made an experimental test of the validity of König's formula, especially from the aspect of the assumptions which underlie that formula. The most important of these are: (1) incompressibility of the fluid, (2) the disc an infinitely thin ellipsoid, (3) absence of viscous forces or spilling of the fluid over the edges of the disc. The first will not matter except when the wave-length of the sound is comparable with the dimensions of the disc, while the correction for the second factor has been given by King, but there was little previous work on the viscous effect. Trying their results on King's corrected formula (which also allows for diffraction of the

incident radiation) they find discrepancies up to 10 per cent, which they attribute to the third factor. (The experimental values of the couple on the disc were determined by some careful repetitions of the original method of König, who placed it inside a box set in vibration with a known amplitude and presumably dragging the air with it.) One cause was found to lie in the departure of the aerial motion from the irrotational form which it takes up in the absence of viscosity and inertia, smoke pictures showing the existence of vortices one at either edge of the disc and on opposite sides of it (cf. figure 2). The critical velocity for the departure from irrotational flow could not, however, be determined, since the vortices were present at all the amplitudes of vibration which could be given to the air in the box. Over the range 10 to 50 c./sec. they suggest that the factor 4/3 which occurs in King's equation for the couple be changed for the empirical factor 1·47 if better agreement with practice be desired, though it seems possible that this factor may vary with amplitude and frequency (i.e. with the energy in the acoustic field) above the range which their experiments cover. Hayashi (1939)

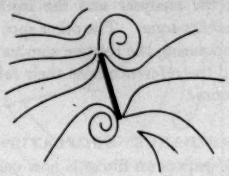

Figure 2. Lines of flow near Rayleigh disc (Merrington and Oatley).

describes a resonant Rayleigh disc for use in water with high-frequency sources; actually he calibrated it with a magneto-striction oscillator of 17·6 kc./sec. fundamental frequency. The same author also uses a thermocouple for the pressure amplitude in sound waves and a hot-wire for velocity in a tube driven by a telephone at one end. He claims the amplitude in each case to be proportional to the thermocouple output (*vide supra*) or the readings of the valve voltmeter in the output of the amplifier which he connects to the hot-wire; also that frequency times output is nearly constant in both instruments; range 100 to 1000 c./sec. Glover and Baumzweiger (1939) describe a "pistonphone" suitable for calibrating Rayleigh discs and like instruments.

Hertz and Mende (1939) discuss the pressure arising at the surface of two liquids irradiated by sound which causes a general movement of the liquid towards the side in which the velocity of sound is the greater. They tried a layer of carbon disulphide under water with the sound source in the former. Although the difference in specific impedance amounts to a mere one per cent, because the velocity is different the energy density is less on the topside of the interface and oil is driven up into the water in the form of a local emulsion. That they had lit on the true cause of the phenomenon was confirmed by the fact that it was equally

vigorous when the energy hit the surface at an acute angle, showing that it was not merely a radiation pressure which lay at the seat of the little fountain. The theory pertinent to this effect was given by Brillouin in 1925. Mendonasse (1939) reiterates this theory and deals with certain new aspects of it. Sacerdote (1939) uses energy density as a parameter in dealing with a number of theoretical problems, e.g. in architectural acoustics, because he finds it is often simpler to use functions of energy density than functions of pressure, as is more often done by theoreticians who have to describe sound fields. Schaefer (1939) shows that for feeble sounds they coincide.

Carrière (1939 b) describes an analyser for measuring the relative phases of the harmonics in a complex note. It is an electric wave-analyser of the type which can be obtained commercially, operating on the beat principle. To this the author attaches a twin-spot cathode-ray oscillograph, in place of the usual micro-ammeter or valve voltmeter in the final stage. One of these spots is made, in conjunction with the time-base, to trace out the evolutions of the beats between the particular harmonic filtered out in the analyser and the injected sinusoidal oscillation with which it beats, as reference tone. The other spot moves under the influence of the latter alone. By comparing the relative stances of the peaks on the screen in the compound note and the reference tone their relative phases are deduced; and so for any other harmonic.

§ 4. ACOUSTIC IMPEDANCE: PROPERTIES OF MATERIALS

Lindsay in his series of papers on filters is now considering the case of compressional waves incident *obliquely* on a stratified medium like a compound partition (Lindsay, 1939 a). The medium is idealised into an infinite series of alternating layers, firstly composed of two different fluids, and then with one solid and one fluid. The latter is a much more complicated problem, since compressional waves incident from the fluid strata will set up both shear and dilatational waves in the solid strata. In the first case the effect of the obliquity is that the cut-off for the first transmission band of frequency rises with the angle of incidence (measured from the surface), and the successive transmission and reflection bands characteristic of a low-pass filter are all displaced towards higher frequencies. For a given angle, increasing the ratio of the thicknesses in the two-fluid case decreases the width of both transmission and reflection bands. A not unexpected selective reflection occurs in the latter type of filter at certain angles of reflection, corresponding to the Bragg law for the reflection of x rays from a crystal lattice. In another paper Lindsay (1939 b) gives a general account of filtration in solid media which includes references to recent developments in theory and practice.

Hall (1939 a) and Trimmer (1939) both deal with the impedance of pipe combinations. Hall's device is an acoustic transmission line, having a telephone source at one end and an impedance to be measured attached to the other. The "line" is an annular groove of rectangular cross-section cut in a brass bed-plate

(cf. figure 3). The top of the groove is closed by a tight-fitting plate which can be rotated and by so doing carries a microphone round the groove. Thus the pressure amplitude from point to point along the line can be plotted by rotating the upper plate (which carries a graduated disc). The line must comprise at least one half wave-length of the tone in question, as the calculation demands the relative pressure amplitudes at a maximum and a minimum. The apparatus shown can be used from 270 c./sec. upwards. The effect of the size of the microphone as an obstacle to the sound field is shown to introduce no error in the results, but it lowers the sensitivity of the apparatus. Trimmer deals theoretically with a number of pipe combinations, most of which are not new. He makes the point that combinations whose impedance would indicate identity of fundamental

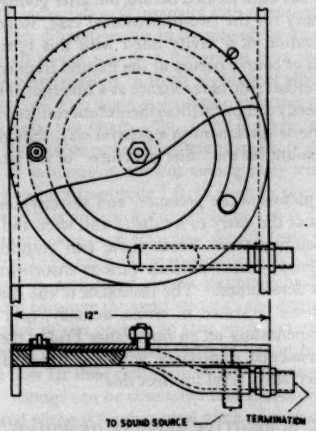

Figure 3. Artificial acoustic line (Hall in *J. Acoust. Soc. Amer.*).

resonance will not necessarily agree in range and strength of harmonics, though there is no way of predicting the harmonic structure from impedance theory unless the characteristics and overtone content of the maintaining source can be specified in a mathematical formula, a thing harder to do in sound than in the corresponding electrical problem.

In two theoretical papers, Morse (1939) deals with a standing-wave system in association with an absorbent wall. In the one case, the waves impinge normally on the absorbent material at one end; applications to the acoustics of rooms are considered in the light of the theory which is worked out for this rather special case. The second case is one which has frequently exercised the mathematical ingenuity of workers in acoustics, i.e., that in which the standing-wave

system is set up *alongside* the absorbent, for instance when it covers one wall of an organ pipe. He shows that, other things being equal, a wall which possesses stiffness (negative reactance), as well as resistance, is more effective as an attenuator of sound than one with mass (positive reactance) plus resistance. Generally, too, a complex wave entering one end of a long conduit will emerge with relatively more fundamental in its make-up than initially. Any acoustic treatment which absorbs sounds in the form of pure tones will be even more satisfactory when complex sounds are in question.

Rogers (1939) and Leedy (1939) attack the problem of the impedance of a vibrating plate backed by an air-space, with a view to the use of such a device as a low-frequency absorber. They also imagine a pair of these as end plates to a pipe. This problem has been tackled before, but after giving a rigorous treatment and graphical summary for the two-dimensional case, Rogers attacks the three-dimensional reverberation of a cavity lined with this type of absorbent. The maximum absorption of course occurs at the natural frequencies of the plate-air-space and falls off on either side of resonance as a function of the damping afforded by the air cavity. Leedy compares these theoretical results with some experiments of A. H. Davis and finds agreement in a general way, though the theory does not seem to be able to account for the " fine structure " of the {absorption, frequency} curve.

Wisotsky (1939) measures the pressure- and velocity-amplitude by means of suitable microphones at the entry to a tube (8 cm. wide and 0·5 cm. thick in the wall) driven by a loud-speaker. From these two amplitudes the impedance looking in from the loud-speaker towards various absorbents and rigid stoppers bored with holes was determined. The resistance of the constrictions of various diameters could then be reckoned in terms, actually, of a calibrated material, that is to say, by compensating on an impedance bridge the pressure amplitude in front of the experimental impedance with the standard, raising the output of the source to the necessary level to effect this.

§ 5. SUBJECTIVE ASPECTS OF SOUND

Last year we described a research on the changes in vowel structure which ensue with changes in the pitch of intensity of their utterance. Now we have an equally important possibility discussed by Black (1939), i.e., does the recorded form of the vowel depend on the consonant which precedes it or are the two sounds quite independent? Experiment favours the second alternative, and this is fortunate for the scientific phoneticians, as it supports their hope of being able to ascribe a particular wave pattern to each standard vowel-sound in whatever circumstances it may first itself. Black states in regard to an earlier research in which wave-to-wave analyses of seven pronunciations of the word "*top*" by a certain speaker were obtained: "It appears that any one of the records could be used for obtaining a description of the speaker's vowel produced under the conditions of the experiment. Despite instability of the vowel and subtle

changes it makes with the course of time, it may be a well-established product for a single individual." To test the possibility that this constancy applies only when the vowel is preceded by *t* and followed by *p*, the same subject uttered the words *top*, *pop*, *fof* and *cock*, using perceptibly the same vowel throughout the series. Analyses of the records revealed that the physical vowel changes very little in company of voiceless consonants, though the centroid of frequency in the vowel shifts somewhat as between one consonantal appendage and another.

Some work by Dunn and Farnsworth (1939) is concerned with the sound field round the head during speech. A speaker in a seated position in an absorbent room simulating open-air conditions repeated a fifteen-second phrase of connected speech while a microphone was traversed through planes across his sound field. Except close to his mouth the frequency spectrum was more or less constant, but there was naturally greater variation of intensity with direction at high frequencies; in fact, above 5000 c./sec. diffraction due to the size of the mouth-aperture became important. The speech power was then integrated over the whole field and its ratio to the value at a point directly in front of, but not close to, the mouth estimated. It is then assumed with some justification that this same ratio will obtain when other patients take the chair. So, from a single measurement at this standard station, the total voice power of different speakers may be derived. The error introduced by an assumption current among early workers in this subject, viz., that the speech pressure amplitude 1 ft. in front of the speaker was the same all over a hemisphere of this radius, works out to be about 2 db.

A number of workers have been remaking speech from the component tones after analysis. Dudley (1939) has built an apparatus, which he calls the "vocoder" in which the synthesis is performed almost instantaneously and automatically, so that the operator can compare the synthetic with the original voice. Modifications can be introduced in the process so that such factors as intonation (pitch variation during speech) can be increased more or less in relation to the original, so that one may note how far their reduction or intensification affects the intelligibility of the speech. Songs can be converted into monotone chants or whispers. Timbre may be altered without fundamental pitch of enunciation and vice versa. "A new method is available for studying the nature of tremulous and pulsating effects such as the vibrato, trill, warble, tremolo and quaver, the method being of especial value because it inherently differentiates the pitch variation from the spectrum variation." Kellogg (1939) plays a more drastic trick on speech by reversing it completely. This is a thing that anyone may try if he can convert his gramophone pick-up into a pusher instead of a tractor, though we do not recommend this to be tried on a valuable record! With a talking film it is easier to arrange. The reversal does not prove severe as far as the intelligibility of diphthongs like the *g* and *i* in *gibe* is concerned. Final consonants being inconspicuous in rapid talking, their inadequacy becomes more noticeable on reversal.

We may at this point describe new methods for analysing fluctuating or transient sounds, since these are of the greatest use in connection with speech

sounds. Brown (1939) uses the diffraction of light from a sound-film for this purpose. A general theory is given of the spectra from such a generalized optical grating in motion through the spectrometer. The diffraction pattern may be regarded as an acoustic spectrum of the original sound averaged over a time T equal to L/V, where L is the width of the gate through which the film passes with speed V; to analyse transients, T must not be too large or too small. Unfortunately, the dispersive power is necessarily of low order, and low dispersion cannot be counteracted by magnification owing to irregularities in the emulsion and the film surface generally. Imahori (1939) employs a photographic device. An analysing plate on which a number of straight dark lines radiate out from the y axis in such a way that the light intensity in the shadow of the plate varies as $a+b \sin 2\pi x/\lambda$ along the x axis, while the wave-length λ varies in linear fashion in the y direction. The incident light's intensity is modulated in synchronism with the sound to be analysed and a photographic film moved at constant velocity

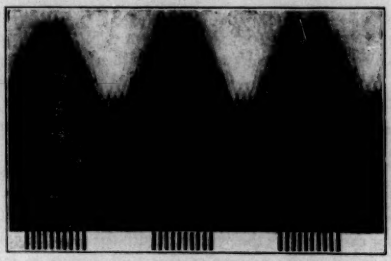

Figure 4. Analysis of square wave by light diffraction (Imahori).
(Reproduced by kind permission from *Nature*.)

along the x direction behind the analysing plate. If the source consists of a harmonic series of vibrations, the film after development shows sharp lines the blackness of which varies sinusoidally over a uniformly brilliant background, one line corresponding to each component. These derive from stroboscopic interference between light fluctuations and motion of the analysing plate relative to the moving film. If the number N of lines inscribed on the plate is not large enough, these lines broaden out into bands whose overlapping limits the resolving power. Fluctuations can be traced by making N sufficiently small, so that the resolving power in respect of frequency is sacrificed to that of time. An example is given (which we reproduce in figure 4) of the analysis of a "square wave" of 10 pulses succeeded by an intermission of 10 periods, then 10 pulses again, and so on; the level n_0 corresponds to the fundamental frequency where the band spacing equals the wave-length of the pulses, whose sequence is illustrated by the pattern below the photograph. The author has also tested his device on speech sounds.

The variation of pitch and the crossing of overtones over the formant of the vowels were clearly recognizable in the analysis.

Passing on now to ***hearing***, we encounter a number of papers dealing with audiometry and threshold measurement. Burr (1939) has carried out such work in a sound-proof chamber at McGill University in which a warble tone for the range above 1000 c./sec. is used. It was established that the pressure in the acoustic field was objectively a replica of the electric power output of the oscillator, and the author considers that such a field is made up of the co-existence of a number of pure tones, with a constant frequency interval equal to the frequency of modulation. These two superposed fields originating in the warble appear to behave as though emitted by a pair of independent sources.

Lifshitz (1939) investigates what he calls the "fluctuation of the hearing threshold". The ear is given impulses of equal duration with a loudness of 40 phons. The series of impulses in groups of 5 is then continued, but the intensity is cut down until a state is reached at which one or other of the impulses gives an impression of shorter duration than its companions. At a lower level one seems to disappear altogether, then two are missed, and so on (in spite of the fact that all have the same objective intensity). When only one is perceived, on the average, in each group of five, this level is rather arbitrarily defined as the hearing threshold, while the domain in which part of the signal miscarries is called the "fluctuation of the threshold"; it begins about 5 to 7 phons above the ultimate threshold. The experiments also show that near the limit the ear can appreciate the duration of a sound more precisely than its loudness. Brittain (1939) prefers a continuous spectrum noise to the usual 1000 c./sec. reference tone for this work, because one does not then have to guard against setting up standing waves in the experimental chamber. It is found that the mean level for the continuous spectrum is 20 db. higher when the two are adjusted to equal loudness. Starting with low intensities, as the level of the noise is raised, nerves from successively greater lengths of the cochlea become activated as the intensity at any pitch reaches the threshold for that particular frequency. Estimates of the loudness of the noise of a refrigerator by a number of listeners were made both in terms of the continuous spectrum and the 1000-cycle tone, using the relationship between the reference sources first established. They agreed quite well.

Von Bekesy (1939) has been measuring the threshold for bone conduction by a piezo-electric manometer. An electrodynamic pistonphone having the quartz cell at the end of the piston (after the fashion of the special piano hammer described in § 2), which abuts upon the skin of the patient, is set in operation at a determined frequency and the pressure amplitude measured when the threshold of hearing (by bone conduction) is attained. The resulting audition curve follows closely that for hearing of air-borne sound. It is then possible to calibrate the tuning fork which is used for tests of bone-conduction by the ear specialist with its stem in contact with the skin of the patient by inserting the quartz manometer in the stem of the fork and comparing its reading in that position with the same threshold when on the pistonphone.

Stevens and Jones (1939) continue their researches on the mechanism of hearing by electrical stimulation—the so-called electrophonic effect—and suggest that the mechanical motion postulated inside the ear to account for stimulation of the sense of hearing has its origin in an electrostatic transducer which they locate in the middle ear, which is supposed to fill the rôle of a condenser receiver. The sound lacks the purity of that heard in the normal way and, further, it is noteworthy that polarization by a direct current distorts the sound more until at a certain value the listener fails to appreciate the fundamental in the A.C., leaving the second harmonic predominant. Since small potentials are always produced at electrode boundaries and in the body of the listener, one might expect that the null point for the fundamental would occur when these potentials just equalled the polarizing voltage, an expectation which is partially borne out by the results. It is assumed that the response follows a second power law, i.e., is proportional to $(a+b\sin\omega t)^2$, in which the fundamental in the expanded form $(2ab\sin\omega t)$ is a function of the D.C. voltage, while the octave $(\frac{1}{2}b^2\cos 2\omega t)$ is independent thereof. An electrostatic transducer fits these requirements nicely, and the small condenser formed by the ear-drum and the opposite wall of the middle ear suggests itself as the seat of the electrostatic potential. If the threshold potentials are known, it should be possible to utilize what is known of the physical dimensions of the ear to calculate the absolute threshold pressure on the ear-drum (as in the similar process of calibrating a condenser microphone). The authors take the average effective pressure on the ear-drum as being equivalent to a pressure on a pair of plates of a condenser with a separation of 1·5 mm., and arrive at a value of $1{\cdot}1\times 10^{-5}$ dyne per sq. cm. for the threshold of the second harmonic.

Wallach (1939) discusses the binaural effect from the aspect of the location of the angular elevation of a source of sound by rotating the head about a *vertical* axis only. If δ is the angular elevation in question measured from the equatorial plane of the head movement, and β is the angular displacement of an axis through the two ears, then the lateral angle ψ, i.e., the angle which the direction of the source makes with this same axis, can be computed from $\sin(90°-\psi)=\sin\beta\cos\delta$ or approximately $\beta.\cos\delta$. A loud-speaker source (unseen) defined the angle ψ in front of and above a listener who turned his head to vary β, pointing up at the angle from which he perceived the sound to come as he did so. The latter angle δ was compared with the corresponding value computed from the formula and fairly good agreement obtained except when δ approached 90° (source overhead), as one would expect.

Littler (1939) determines the natural frequency of the air in the external ear (*meatus*) which in the human being lies in the neighbourhood of 2000 to 3000 c./sec. and to a certain extent accounts for the superior sensitivity of the ear in this part of the gamut. His method is to replace the air in this cavity by hydrogen, so raising the resonant frequency in the ratio of the velocities of sound in the respective gases, and to re-determine the threshold in this pitch region. At 2500 c./sec. the reduction in intensity of a perceived sound is about 12 db., decreasing to zero

at 6000 c./sec.; above this the introduction of hydrogen raises the level, e.g. by 9 db. at 9000 c./sec. Higher still the air audibility is again attained. Meatus resonance may at the most favourable pitch produce a threefold amplification of sounds impinging on the ear-drum. Fleming (1939) criticizes these experiments on the ground that they were made with the outer end of the meatus closed by a telephone diaphragm, whereas normal hearing, of course, leaves the end open. Comparisons aimed at testing the gain of hearing aids—which is the aspect that Littler is interested in—should therefore take as standard the conditions of unassisted hearing with open-ended meatus.

§ 6. TECHNICAL APPLICATIONS OF SOUND

Microphones and loud-speakers. Massa (1939) considers the directional response of pressure-gradient microphones in relation to their size. If the pressure element is terminated by an obstacle of finite size, the effect of the obstacle is to increase the acoustic path between a plane wave-front and the pressure-actuated element. The phase shift increases with the frequency, and consequently the directional response departs from the theoretical cardioid. This in turn raises the response of a unidirectional microphone in the way that is not wanted, i.e., from behind. Glover and Baumzweiger (1939) have a moving-coil pistonphone suitable for the measurement of pressure in sound fields, while Mason and Marshall (1939) have a tubular directional microphone used in broadcasting events wherein conditions make it necessary to erect the microphone at a distance, for example, at parades and pageants.

Barrow (1939) has constructed a multitone generator which should prove useful as a sound source for certain acoustic measurements. It is made up of neon-tube oscillators or thyratrons employing shock excitation with a multiple commutator driven by a synchronous motor used to switch in the grid circuits and so secure exact frequency and phase relationships between the component oscillators. The final generator includes eleven of these simple oscillator circuits. These standard tones are then recorded on gramophone discs at a speed which gives a frequency of 500 c./sec. to the fundamental. From this any desired frequency may be obtained by playing back at a suitable speed.

Reverberation in buildings. Hall (1939 b) describes a convenient reverberation-time meter in which during the process two relays are tripped in turn when the sound level passes through two pre-set values. The first starts a condenser charging up through a resistance, the second stops the charging by removing the applied voltage. The potential then remaining across the condenser plates is applied to the grid of a valve. Since this potential depends on the time between the two trips, the milliammeter in the anode circuit may be graduated directly in seconds of reverberation time. Morrical (1939 a) has a useful graduated protractor on which reverberation times can be read from the slope of the {log intensity, time} curves of the decay of sound in a room.

Hunt (1939) proposes to use the steady-state characteristics of a source and

receiver system in a reverberant chamber in place of the transient analysis given by apparatus of the class just named. If it be objected that in practice the frequency and power of the source vary so rapidly that transmission seldom approximates to a steady state, it is pointed out that, as in electrical networks, designation of the terminal impedances and propagation constants of the system is enough to determine both the steady-state and transient characteristics of the transmission path which, in the present application, denotes the acoustical features of the environment. These characteristics in turn will be made up, according to recent conceptions (*vide supra*, § 4) of a certain admixture of the normal modes of vibration of the air in the room. Hunt carried out experiments in a model reverberation chamber whose linear dimensions could be varied and were of the order of 2 ft., giving a number of discrete and well resolved natural frequencies lying between 250 and 1500 c./sec. The source is "injected" into the chamber at a particular point via a narrow tube, in order not to upset unduly the pattern of standing waves by obtruding an out-of-scale loud-speaker. To avoid resonance in the supply tube, it was packed with wires and the sound from an external loud-speaker passed along the interstices. The impedance of such a nest of tubes is so high that the open end may be taken to represent a constant-volume velocity source. Typical {response, frequency} curves were made with a microphone at a set position relative to this source, both with and without a lining of absorbent material on the walls of the model room. The curves are naturally much indented by the normal modes of the model. It is suggested that absorption coefficients may be calculated from the relative breadth of these resonance peaks with the aid of the formula

$$\alpha = 1 - \left(\frac{Z\cos\theta - \rho c}{Z\cos\theta + \rho c}\right)^2,$$

Z being the real part of the impedance at the surface for waves incident at an angle θ while ρc is the specific impedance of the air in the chamber. In the preliminary work, a value of Z has been chosen to fit the plotted points on a graph of α against θ. The work is in progress. Other aspects of such methods are discussed in a Symposium on the Measurement of Absorption Coefficients and the Application of Acoustical Absorbents held at the Tenth Anniversary Meeting of the Acoustical Society of America (1939), in which Morse, Bhatt, Beranek, Bolt and Maa took part. The last-named pair of authors have calculated the normal modes of vibration of a three-dimensional rectangular enclosure at greater length in separate papers. Brillouin (1939) also enters this field by calculating the modes of forced vibration of a space limited by absorbing surfaces. Non-uniform plane waves of special structure are set up with elliptical vibrations in a plane containing the direction of propagation having amplitudes following an exponential law and with phase-velocity and velocity of energy-propagation respectively equal to and less than the velocity of sound.

Skudrzyk (1939) has a similar paper for the case of non-planar walls and diffuse reflection. The latter factor he first brings into his work in the form of

conical stoppers to tubes so placed as to make the reflected sound ricochet from wall to wall of the tube. He supports results derived from impedance considerations by experiments in which the tube is driven by a loud-speaker at the opposite end to the stopper while a microphone is applied to various points in the side walls. The field in the tube naturally becomes very complex at frequencies above the lowest radial resonance. To get diffuse reflection of a less specialized kind, he builds up an apparatus not unlike that used in optics in the estimation of glare from an illuminated surface. Radiation from a magneto-strictive source falls upon the surface at a given angle, while an exploring microphone on an arm centred at the point of incidence explores the reflected field in the plane of incidence. Results for substances like macadam stone about 12 to 18 wave-lengths wide showed, however, little similarity to optical glare. This is not surprising when one marks the difference in scale of the two phenomena. Using a supersonic source and a rough-cast surface having rugosities of the order of 10 to 15 wave-lengths in size, on the other hand, results consonant with the laws of geometrical optics were obtained. The obliquity law of Lambert, for instance, satisfies its acoustical analogue.

Other papers in the American symposium which deal more explicitly with the measurement of absorption coefficients are to be found under the names of Hunt, P. Sabine, Stanton and Willig. The last-named (Willig, 1939) in an earlier paper suggests means by which a value for the average absorption offered by a specimen may be derived from outdoor or tube methods operated at isolated frequencies. Meyer and Schoch (1939) are also concerned in this matter, though rather from the opposite aspect. As a result of their work they suggest that when the acoustic properties of a specimen are being examined from the point of view of its usefulness in an auditorium, firstly, that the reverberation time should be determined with the smallest steps in frequency that are reasonably possible, covering the whole gamut; secondly, that tests of the material ought to be made on three walls separately, these not being parallel, and in comparatively small areas.

Sound insulation. Morrical (1939 b) also objects to the employment of average figures for sound transmission on the ground that any average, however determined, does not afford a true picture of the action of an insulating partition. It does not, for instance, reveal a pronounced weakness at some point in the pitch range or pronounced attenuation at frequencies where it is not of practical importance. A better approach to the specification of the effectiveness of a sound-deadening structure is to consider the loudness level to which the partition will reduce the sounds most often encountered in practice. The author exhibits ideal contours of transmission loss against frequency, necessary to reduce music or speech to a loudness level specified on each contour, and compares these with the corresponding contours applying to actual partitions. Constable (1939) made an investigation of the relative importance for transmission of the paths by which sound can pass through the structure of a building (a block of flats) from a room

on one floor to other rooms on the same floor and on the one below. Not only the intervening partitions but the flanking walls and floors proved to be conductors for the sound, indirect paths such as these latter becoming responsible for a good share in the transmission in certain cases. In most buildings approximate equality of weights of walls and floors makes the amplitudes of vibration of all of them to be of the same order, and this fact fortunately limits the importance of indirect routes. Since the area of flanking walls in rooms of average shape tends to be large in comparison with the lighter partitions, their contribution is likely to exceed that of the partitions; and this proved so. Without additional weight the partition could not restrict the vibrations of the flanking walls to which it was attached, so that there would be no point in making the partitions much more insulating than the heaviest wall to which they might be pinned, which means that they cannot, in average buildings, be given a useful reduction factor of more than 60 db.

Noise. Robinson (1939) has given an account, with more detail than any we have yet encountered, of alleviation work in connection with noise on board ship. He shows what success has been achieved in the engine room and points out that the use of magnesia has enabled the engineer to combine acoustic with thermal treatment in the one material. Other treatments with cork slabs and sprayed asbestos have met with success, the latter being especially useful in treating ventilator conduits. The "singing propeller" nuisance, to which we have referred in previous reports, has recently engaged the attention of Shannon, Conn and Arnold (1939), who, after recapitulating the salient features of the phenomenon, proceed to consider the fundamental causes of propeller noise in relation to the hydrodynamic forces on the aerofoil sections of which the blades are composed. They opine that the fundamental cause of singing is wake variation over the propeller disc; this can be kept down by avoiding all obstructions to the flow and setting the propeller well aft, clear of stern-post and rudder. The form of the leading edge of the outer third of the blade is important in controlling the risk of breakdown in the flow, of cavitation and of vibration. It is best to make it sharp and free from sudden reversals of curvature.

Obata and Yosida (1939) draw attention to the intense combination tones produced in the flutter of an airscrew blade (which indeed may be a contributory cause of the water noise just mentioned). Records of the noise of an airscrew both below and above the critical speed at which flutter sets in—about 1800 rev./min.—have been made. It is held that many of the tones found in the noise spectrum at super-critical speeds are combination tones formed between the "sound of rotation" (due to the pressure difference fore and aft of the blade) and one or other of the torsional vibrations of the blades themselves (which may involve twists of the order of 5°, corresponding to displacements of several cm. of the edges of the blades). The objective existence of these compound notes was shown by resonators tuned to expected frequencies and excited by pairs of loudspeakers set to produce intense pure tones of frequency equal to the two suspected components in the noise spectra.

Electro-acoustics. A discussion on this subject, held by the Physical Society (1939) under the leadership of Dr. Drysdale, ranged over the following subjects: ideal receivers and absorbent materials; the quality of reproduction and its testing; improvements in telephone receivers; the resistive and reactive components in the impedance of materials. It is not possible to summarize these papers in a reasonable space, but one point made by several of the speakers is worthy of emphasis, viz., that the ear must be the final arbiter in all acoustic testing.

REFERENCES

Acoustical Society of America, 1939. *J. Acoust. Soc. Amer.* **11**, 37.
Alleman, R. S., 1939. *Phys. Rev.* **55**, 87.
Arts, J., 1939. *J. Acoust. Soc. Amer.* **10**, 327.
Auger, L., 1939. *C.R. Acad. Sci., Paris*, **209**, 508.
Barrow, W. L., 1939. *J. Acoust. Soc. Amer.* **10**, 275.
Becker, H. E. R., 1939. *Ann. Phys., Lpz.*, **36**, 589.
Bekesy, G. von, 1939. *Akust. Z.* **4**, 113.
Bhagavantam, S. and Rao, C. V. J., 1939. *Proc. Ind. Acad. Sci., Bangalore*, **9**, 312.
Biquard, P., 1939. *Rev. Acoust.* **8**, 130.
Black, J. W., 1939. *J. Acoust. Soc. Amer.* **10**, 203.
Brillouin, J., 1939. *J. Phys. Radium*, **10**, 497.
Brittain, F. H., 1939. *J. Acoust. Soc. Amer.* **11**, 113.
Brown, D., 1939. *Proc. Phys. Soc.* **51**, 244.
Burr, E. G., 1939. *Trans. Roy. Soc. Can.* **33**, 73.
Carrière, Z., 1939 a. *Rev. Acoust.* **13**, 12.
Carrière, Z., 1939 b. *J. de Phys.* **10**, 14.
Cerovska, Jarmila, 1939. *J. Phys. Radium*, **10**, 97.
Colwell, R. C., Stewart, J. K. and Friend, A. W., 1939. *Phil. Mag.* **27**, 123; *J. Acoust. Soc. Amer.* **11**, 147.
Constable, J. E. R., 1939. *Proc. Phys. Soc.* **51**, 53.
Davy, N., Littlewood, J. H. and McCay, M., 1939. *Phil. Mag.* **27**, 133.
Dudley, H., 1939. *Proc. Nat. Acad. Sci., Wash.*, **25**, 379; *J. Acoust. Soc. Amer.* **11**, 169.
Dunn, H. K., 1939. *Rev. Sci. Instrum.* **10**, 362.
Dunn, H. K. and Farnsworth, D. W., 1939. *J. Acoust. Soc. Amer.* **10**, 184.
Dwyer, Ri J., 1939. *J. Chem. Phys.* **7**, 40.
Field, G. S., 1939 a. *Canad. J. Res.* **17**, 197.
Field, G. S., 1939 b. *Canad. J. Res.* **17**, 141.
Fleming, N., 1939. *Nature, Lond.*, **143**, 642.
Fox, F. E. and Rock, G. D., 1939. *Rev. Sci. Instrum.* **10**, 345.
Freundlich, H. and Gillings, D. W., 1939. *Trans. Faraday Soc.* **35**, 319.
Ghosh, M., 1939. *Indian J. Phys.* **13**, 23 and 277.
Glover, R. P. and Baumzweiger, B., 1939. *J. Acoust. Soc. Amer.* **10**, 200.
Groenewald, H. J., 1939. *Physica*, **6**, 303.
Hall, W. M., 1939 a. *J. Acoust. Soc. Amer.* **11**, 14.
Hall, W. M., 1939 b. *J. Acoust. Soc. Amer.* **10**, 302.
Hayashi, T., 1939. *Electrotech. J.* **3**, 103, 175.
Hertz, G. and Mende, H., 1939. *Z. Phys.* **114**, 354.
Hopwood, L. F., Salaman, M. H. and McFarlane, A. S., 1939. *Nature, Lond.*, **144**, 377.
Hubbard, J. C. and Zartman, I. F., 1939. *Rev. Sci. Instrum.* **10**, 382.
Hunt, E. V., 1939. *J. Acoust. Soc. Amer.* **10**, 216.
Imahori, K., 1939. *Nature, Lond.*, **144**, 708
Itterbeek, A. van, Bruyn, P. de and Mariens, P., 1939. *Physica*, **6**, 511.

JATKAR, S. K. K., 1939. *J. Indian Inst. Sci.* **22** A, 1, 24, 39, 59, 93.
JONES, A. T., 1939. *J. Acoust. Soc. Amer.* **10**, 167.
KELLOGG, E. W., 1939. *J. Acoust. Soc. Amer.* **10**, 324.
KNESER, H. O., 1939. *Ann. Phys., Lpz.*, **34**, 665.
KRISHNAN, K. G., 1939. *Proc. Indian Acad. Sci., Bangalore*, **9**, 382.
LEEDY, H. A., 1939. *J. Acoust. Soc. Amer.* **10**, 288.
LIEPMAN, H. W., 1939. *Helv. phys. Acta*, **12**, 421.
LIFSHITZ, F., 1939. *J. Acoust. Soc. Amer.* **11**, 118.
LINDSAY, R. B., 1939 a. *J. Acoust. Soc. Amer.* **11**, 178.
LINDSAY, R. B., 1939 b. *J. Appl. Phys.* **10**, 680.
LITTLER, T. S., 1939. *Nature, Lond.*, **143**, 118.
MASON, W. P. and MARSHALL, R. N., 1939. *J. Acoust. Soc. Amer.* **10**, 206.
MASSA, F., 1939. *J. Acoust. Soc. Amer.* **10**, 173.
MATOSSI, F., 1939. *Phys. Z.* **40**, 294.
MENDONASSE, J., 1939. *C.R. Acad. Sci., Paris*, **208**, 1977.
MERRINGTON, A. C. and OATLEY, C. W., 1939. *Proc. Roy. Soc.* A, **171**, 505.
MEYER, E. and TAMM, K., 1939. *Akust. Z.* **4**, 145.
MEYER, E. and SCHOCH, A., 1939. *Akust. Z.* **4**, 53.
MOKHTAR, M., 1939. *Phil. Mag.* **27**, 195.
MORRICAL, K. C., 1939 a. *J. Acoust. Soc. Amer.* **10**, 300.
MORRICAL, K. C., 1939 b. *J. Acoust. Soc. Amer.* **11**, 211.
MORSE, P. M., 1939. *J. Acoust. Soc. Amer.* **11**, 56 and 205.
OBATA, J. and YOSIDA, Y., 1939. *Proc. Phys.-Math. Soc. Japan*, **21**, 588.
PHELPS, W. D., 1939. *J. Acoust. Soc. Amer.* **11**, 219.
PHYSICAL SOCIETY, 1939. *Proc. Phys. Soc.* **51**, 359.
PIELEMEIER, W. H., 1939. *J. Acoust. Soc. Amer.* **10**, 313.
POHLMAN, R., 1939. *Phys. Z.* **40**, 159.
RAILSTON, W., 1939. *J. Acoust. Soc. Amer.* **11**, 107.
RICHARDSON, E. G., 1939. *Nature, Lond.*, **143**, 638.
RIECKMAN, P., 1939. *Phys. Z.* **40**, 582.
ROBINSON, R. S., 1939. *Trans. Inst. Marine Eng.* **50**, 1.
ROGERS, R., 1939. *J. Acoust. Soc. Amer.* **10**, 280.
SACERDOTE, G., 1939. *Pont. Acad. Sci. Acta*, **3**, 47.
SALCEANU, C., 1939. *C.R. Acad. Sci., Paris*, **208**, 83.
SATA, N. and NAMU, N., 1939. *Kolloidzschr.* **86**, 102; **87**, 186; **88**, 182.
SCHAEFER, C., 1939. *Ann. Phys., Lpz.*, **35**, 473.
SCHMID, G., ROMMEL, O. and ROLL, A., 1939. *Z. Elektrochem.* **45**, 659 and 659.
SCHREUER, E., 1939. *Akust. Z.* **4**, 215.
SCHULZE, R., 1939. *Ann. Phys., Lpz.*, **34**, 41.
SHANKLAND, R. S. and COLTMAN, J. W., 1939. *J. Acoust. Soc. Amer.* **10**, 161.
SHANNON, J. F., CONN, J. F. C. and ARNOLD, R. N., 1939. *J. Inst. Eng. Scotland*, **81**.
SKUDRZYK, E., 1939. *Akust. Z.* **4**, 172.
STEVENS, S. S. and JONES, R. C., 1939. *J. Acoust. Soc. Amer.* **10**, 261.
THOMPSON, L. and RIFFOLT, N., 1939. *J. Acoust. Soc. Amer.* **11**, 233 and 245.
THOMSON, W. T., 1939. *J. Acoust. Soc. Amer.* **11**, 198.
TRIMMER, J. D., 1939. *J. Acoust. Soc. Amer.* **11**, 129.
WALLACH, H., 1939. *J. Acoust. Soc. Amer.* **10**, 270.
WALLER, MARY D., 1939. *Proc. Phys. Soc.* **51**, 831.
WILLIG, F. J., 1939. *J. Acoust. Soc. Amer.* **10**, 293.
WISOTSKY, W., 1939. *Hochfrequenztech. u. Elektroakust.* **53**, 97.
ZACHOVAL, L., 1939. *J. Phys. Radium*, **10**, 350.

MUSICAL ACOUSTICS

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.,
King's College, Newcastle-upon-Tyne

In recent years no more noteworthy instance of the invasion of the domain of the arts by science is to be found than in the part which modern physical research has played in the design and timbre-analysis of musical instruments and in setting the acoustical aspects of music generally on a quantitative scientific basis. Some regret this intrusion, holding fast to the idea that the artist should not be disturbed in his creative activity by mundane consideration of vibration numbers and decibels; yet even the virtuoso has to tune his intrument or get someone to tune it for him, and he could scarcely refuse the gift of a more potent and manageable instrument just because its construction had involved certain processes in the very material laboratories of the acoustic or electrical engineer.

The purpose of this article is to give the general reader some account of recent advances in this sense.

§1. THE FORMANT THEORY

It is unnecessary perhaps at the outset to dwell on the nature of tone-quality or timbre. Most musicians nowadays are aware that this elusive characteristic has given up its secret, and that what the scientist calls the "note" of a musical instrument can be split up into a number of simple "tones", and that it is the relative number and magnitude of these components that determines timbre; tells us, for example, why a violin and a clarinet, or even two violins, do not, when played severally, sound the same when producing a note at the same fundamental pitch. In many cases these component tones are members of the harmonic series, i.e., their frequencies bear simple ratios to the fundamental or component of lowest pitch; but again other tones—not harmonics—often intrude, whose origin is to be traced to the vibrations of the invariant parts of the instrument such as the belly of the violin or the metal tube of the trumpet, and which do not alter much when the player passes from one note to another. These invariant tones make up the "formant "of the instrument.

Apparatus for delineating such wave forms was developed by Miller in the days before the thermionic valve. His phonodeik was certainly the most highly-developed wave delineator of the mechanical class. Now we have condenser microphones, amplifiers and cathode-ray oscillographs to do the work for us, or there is the alternative of the sound film. All these recorders have to be calibrated. Their characteristic {response, frequency} curve must be known and the

resulting wave-forms corrected, if necessary, for non-linearity of response in the analyser. The most convenient apparatus for this purpose is an electric wave-analyser. This electric wave-analyser is like a wireless receiving set in principle, but it covers the gamut of audible sonic waves instead of the supersonic waves which are used to "carry" broadcast music. One proceeds to tune the analyser through its range of frequency from 20 to 20,000 vibrations/sec., picking up the various harmonics of the wave-form and noting the readings of the voltmeter which records the intensity of each constituent signal. (This instrument gives, in fact, an indication like that often given on a wireless set to indicate when the maximum of tuning-in to a station is reached.) The shape of the condenser plates favours the use of a logarithmic scale of frequency, advocated by Young (1939).

If the sound to be analysed is a musical sound of constant pitch and intensity, the current from the recording microphone is taken directly after amplification to the wave analyser without the intervention of the film, but if we are dealing with speech or sung sounds for phonetic studies we have to make a film and analyse a short portion of it, for the voice is never constant enough to ensure exact repetition of the wave form from period to period.

As illustrations of the application of the physical analysis to musical sounds, we shall discuss the acoustics of the violin and the human voice, since a fair amount of work has been done recently in these two fields.

The violin

Both Meinel (1939) and Saunders (1940) have been studying the characteristics of violins in the form of their {amplitude, frequency} curves when played either by mechanical means or in the hands of virtuosi. They hope by comparing sets of such curves to be able to define the physical criteria which mark out the excellent from the mediocre instruments. To Meinel, one criterion seems to be relatively low amplitudes in the upper treble. Also in the timbre the ratio of the amplitude of the fundamental to that of overtones lying above 3000 c./sec. is larger in good violins than in those of poor quality. The tone of a good violin also carries better, so that differences in quality are more noticeable when the listener is at a distance from the player than when he is close up. As the thickness of the wood is diminished, the natural frequencies sink, so that by this cause alone the fundamental may increase in weight relative to the overtones. He ascribes modern failure to surpass or equal the Italian makers of the seventeenth century to the use of too thick wood.

Saunders utters the *caveat* that such old masters as have survived cannot have come down to us unchanged. The rise in musical pitch in the past three centuries (to which we refer in a later paragraph) has made it necessary to strengthen the framework (by thickening and lengthening the bass bar which is glued under the belly), or this expedient may have been employed to produce a louder tone. In some cases this change has actually improved upon the original instrument.

Using a mechanical player, records are made of thirteen tones at intervals of a semitone on each of the four strings, with an additional octave on the high E string. The response curve (cf. figure 1) contains in compact form the summary of all the analyses of these 64 notes. All but one of the principal peaks may be explained as being due to body resonance. The exception is one near the note of the open D string, which is due to the air inside the violin. Saunders's opinion with regard to excellence in quality is that the best violins have a fairly uniform response over most of the range, with an increase near 2000 and a sharp drop above 5000 c./sec. Curves of this type are obtained from both new and old examples. On our figure we show that of a Stradivarius paired with a modern Wagstrom, and a Maggini with a modern Goss. If judgement is based on tone quality as measured in this way, we must, he concludes, confess that modern violins can be produced which

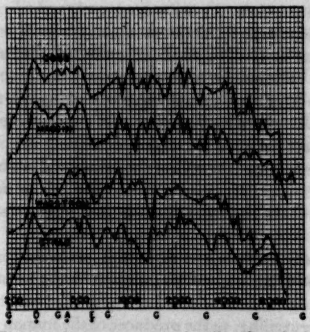

Figure 1. Response curves of violins (Saunders).

copy the characteristics of the old masters. On the other hand, it is often found that less power has to be expended on the old violins to make them "speak", and this is of importance to the player, especially when he is executing rapid passages. He also points out that the tone quality may be profoundly affected by small adjustments, of which the positions of sound-post and bridge are the most important.

The singing voice

The human vocal apparatus consists of the chest (bellows), larynx containing the vocal cords (source of vibration), and the throat, mouth and nasal cavities (resonators). The production of tone has often been likened to that in a reed pipe of the wood-wind class; but perhaps a brass instrument of the trombone type gives a closer analogy, since both the size of the resonators and the tension and length of the vibrating tissues are adjustable. The true or lower vocal cords are set in vibration by approximating their edges under the tension necessary to produce a note of the required frequency, and the breath expelled through the

narrow slit between them. This process sets the edges in vibration either as a whole (normal register) or in part (falsetto register). These vibrations may be seen with a combination of the laryngoscope and stroboscope. Experiments on the excised larynxes of baboons, etc., have shown that, as would be expected, the quality of the unaided vocal chords is raucous and unmusical. It is mainly by the modulating influence of the resonating cavities upon the laryngeal tone that man is able to transcend the lower animals in speech and song, though the parrot and certain members of the ape family possess this faculty to a limited extent.

The vocal cords act as a sort of valve. The pressure of the breath under the closed slit formed by them when they are approximated forces the slit open for a moment, but the increased tension thus set into action causes the cords to slip back and close the slit, which is again re-opened by an accumulation of pressure underneath. In a sense, then, the maintenance of cord vibrations by the one-way flow of breath is analogous to the maintenance of the violin-string vibration by the one-way motion of the bow. The smaller the volume of breath expired per second in singing, the more efficient the maintenance of vibration. A "breathy" voice production involves waste of energy beside leading to throatiness, which is actually caused by the excessive production of high-pitched fricative sounds as the stream of air passes between the clenched cords.

It has been suggested that the rôle of the false vocal cords (two bands of tissue lying above the true vocal cords) is similar to that of the upper lip of the (flute) organ-pipe or of the sharp rim in the mouthpiece of the trombone where it joins the main tube of the instrument—i.e. that an "edge-tone" is produced between the two pairs of vocal cords. Just as in this orchestral instrument, from the *congeries* of vibrations of all frequencies imposed on the breath emerging from the lips, that particular edge-tone is selected and amplified which is appropriate to the physical conditions prevailing, so the producer of the human voice is able to purify the cord-tone by a suitable adjustment of the relative position of the two sets of vocal cords. This part of the vocal apparatus is most closely copied by those decoys or bird-calls consisting of a flattened metal capsule having two circular holes facing each other, through which one blows. We turn to the important modifications produced on this still rather unmelodious "buzzer" note by the resonance columns which it now has to pass through. The natural vibrations of the latter in the absence of a buzzer note may be heard in whispered speech or in whistling, when the vocal bands are drawn aside and a stream of air passes through, practically uninterrupted at the larynx. Actually the damping is very large, much more so than in a wind instrument, owing to the resilient nature of the walls of the cavities. This has the practical effect of broadening all the resonances. It makes, in fact, the testing of all the natural resonances of the system held in a given position by such methods as bringing tuning forks in front of the mouth rather difficult. Damping tends, in fact, to confound the theories of Helmholtz and Hermann on the subject of the wave-forms of vowels.

While it is true that we find, with the latter, that as one sings a vowel up the

scale each cord-note has the formant of the fixed resonance cavities impressed upon it, these resonances are so broad that the probability will be that a number of harmonics in the cord-note will elicit some resonance in the mouth-cavities, giving apparent agreement with the Helmholtz relative-pitch theory of vowel overtones. This has the result that within limits the wave-form of a vowel is preserved as the cord-note is changed. This would not be so in a brass instrument, where rigid walls and consequent sharp resonances impress a very pronounced formant on each note, altering the wave-form as the player moves up the scale.

Turning now to the resonators, we observe that some of these, viz. the nasal cavity and sinuses, are fixed, but that the larger ones, viz. the lungs, pharynx and mouth, are more or less adjustable in size and shape. It is generally agreed that vowels are characterized by one or two resonances of fixed pitch. The mouth may be "set" for these in various ways. Some people tend to speak with their mouths nearly shut. They must compensate for the small change of opening which they allow themselves by large changes of the size of the pharynx when they wish to produce different vowels. Ventriloquists are, of course, adept in this art. On the other hand, a singer may vary the size of the mouth-opening considerably between different vowels; in that case he need not use the pharynx muscles to such a large extent. The best method is doubtless to adopt a fairly wide-open (but not fixed) mouth-opening with suitable variations in the size of the pharynx for different vowels. This will not always be physically possible: indeed, it is impossible to sing high-pitched notes to certain vowels. It is obvious that if a vowel requires a resonance fairly low in the clef, this resonance will not be excited when the vocal-cord note lies high; on the other hand, a vowel like *ee* in *reed*, with high resonances, is readily excited at a high pitch: in fact most sopranos have to make an *ee* do for any vowel at the top of their range.

§ 2. IMPEDANCE PRINCIPLES AND THE ACOUSTICS OF WIND INSTRUMENTS

By adopting ideas from alternating-current theory, the acoustician has been able to solve a number of analogous problems in sound wherever the acoustic elements could be grouped into "inertances" and "capacitances" and built up into circuits whose electrical counterparts had already been solved or were amenable to solution. In the case of an acoustic radiator whose frequency is to be varied over a certain range, the system is assumed to be driven at resonance and the damping may be neglected. In most wind instruments the variation is a step-by-step one, each step being a semitone.

The problem is then usually set in the form:—What adjustments of impedance of the component parts are necessary for this variation? And of the alternatives, if any, which ones will allow the player to progress with the greatest facility from one note to another? For those who are familiar with loud-speaker practice, and particularly that of the exponential horn, it is important to realise that the function of a wind instrument is not primarily to radiate sound in a specified

direction, but to produce a note of frequency determined by the resonant state of the system of impedances of which it is made up. Though exponential horns have been employed on the brass-wind, they cannot be given too pronounced a flare, or their resonance peaks will overlap so much that the player is sore put to it to strike one note without warbling into another—a state of affairs which is closely approached on the modern French horn, to the anxiety of the tyro!

Though the notes on such an instrument are restricted to the harmonic series based on a few set speaking lengths, the wood-wind adopt a different principle. This is to make use of side-holes, whose opening and closing varies the arrangement and value of the impedances so linked in the instruments. A simple wind instrument like the flageolet is, in fact, the acoustical analogue of a short artificial electric line, with the side-holes set as earthed branches to the main conduit. The "conductivity" (or "admittance", to use a modern term) of these orifices can be nullified by covering with a finger or key—or reduced by half-covering, a practice adopted by flautists of the older schools. The flute then is the analogue of the cable (the air in the tube) loaded by inductances (the side-holes). Thus, if all holes are covered except one, the system may be treated as a cable (up to the site of the open hole), terminated by an inductance and another length of cable in parallel, and the net impedance equated to zero for resonance. (Suitably placed vent-holes of smaller diameter than the note-holes assist the player to produce harmonics). When several holes, remote from the mouthpiece, are open, the contribution of those lower down may be neglected, but not that of those near the first open hole. Thus the note formed by covering the holes 1, 2, 3 and 5 may be intermediate in pitch between that fingered 1, 2, 3, 4 and that of 1, 2, 3, 4, 5.

In such ways the scientist is enabled to predict for the instrument maker the location and size of these note-holes, which were formerly adjusted by the method of trial and error. Much yet remains to be done in the precise estimation of "conductivity", however, a matter to which we have referred in previous *Reports*. Thus the conductivity of a given orifice is not the same whether it terminates a tube or whether it is at the side, or again, when it divides the tube as a constriction.

§3. STANDARDS OF PITCH

The infinite local variations in tuning pitch began to settle down to some measure of uniformity in the early part of the nineteenth century. A conference of physicists at Stuttgart in 1834 proposed that the A for tuning should be set at 440 c./sec. This was endorsed for England by the Society of Arts in 1850.

Unfortunately, the 1850 recommendation remained little more than a record on paper, and it appears, looking back, that the Society itself was partly responsible for this state of affairs through an unfortunate ambiguity in its proposals. At that time it was common to use a C tuning-fork for determining instrumental pitch. Now if you say that A shall be 440, it is natural to determine the C above from its relation in just temperament, which is a minor third ($=6/5$) higher. It is, however, well known that the exigencies of playing on keyboard instruments demand a

modification of just temperament (Williamson, 1939). On the pianoforte, then, the R.S.A. standard would require C = 524 instead of C = 528 of the orchestra. Standard tuning forks were for a time adopted in *orchestral* tuning (in just temperament) corresponding to a pitch for A, if it had persisted, of 444 instead of the 440 intended by the Committee.

The situation became complicated when a French Commission containing such well-known names as those of Berlioz, Meyerbeer and Rossini decided in 1858 for A = 435, and entrusted to Lissajous the construction of a standard fork of this frequency.

Another committee was appointed, and in 1886 reported in favour of giving up what had become known as the Society of Arts pitch (440) in favour of the French (435): "The Committee think it would be a pity to cause a schism for the sake of a few vibrations and . . . recommend that the influence of the Society should be employed in furthering the adoption of the French pitch."

In spite of this, the tuning pitch was made higher and higher, and at the close of the century attention was again directed to standardizing the pitch at a lower value, where it would be more within the compass of singers.

Covent Garden Opera was lowered in 1880 to French pitch to please Patti, who, as she got older, was unable to sing soprano parts at the old high pitch, and a number of choral societies followed. The churches were tied to their organs until these were scaled down. The fact that in many instances the lowering involved was approximately a semi-tone aided the organ conversion, as the whole stop of pipes could be moved up one position in each rank, the top one discarded and a new bottom pipe supplied at a moderate cost. In 1896, the Philharmonic Society adopted 439, and this pitch was endorsed by the pianoforte trade.

The army was the last stronghold of high pitch, but their pitch was lowered in 1927, so that —with the exception of a few local choral societies and orchestras— the pitch of musical performance hovered round A = 440, within a few vibrations, over the whole of Europe at any rate, as tests of broadcast performances within the last few years have shown.

To bring this account up to date it should be added that in 1938 a sub-section of the Acoustical Terms Committee of the British Standards Institution was appointed from among British scientists, professional musicians and instrument makers, together with representatives of the B.B.C., and decided to re-inforce the original suggestion (A = 440) of a century earlier. This was put to an international gathering of similar interests in 1939, unanimously agreed upon and endorsed by the British Standards Institution, London, from whom a pamphlet (British Standards Institution, 1939) on the matter can be obtained (Kaye, 1938; Kaye, 1939; Richardson, 1940).

Van der Pol and Addink (1939), for example, have made an interesting series of measurements of the variations in pitch between different orchestras and in the same orchestra during a performance. To compare the pitch of the tuning-

note it is sufficient to compare the pitch during tuning with the standard 440 by counting beats or by the use of a wave-meter (*vide infra*). In order to assess the variation in pitch during performance, the music from a wireless receiver was passed through a filter circuit which limited the electrical vibrations passing to those in the vicinity of A—actually from A♭ to A♯—and the frequency of such vibrations as passed the filter compared with the standard.

The mean of A for each country observed was within 1 or 2 cycles of the standard, but in some the tolerance was greater than in others. For example, of 150 British broadcasts the mean was 439, the minimum 433 and the maximum 442; of 106 Dutch broadcasts, the mean was also 439, but with a minimum of 430 and a maximum of 447. Very interesting are their observations on the change of pitch during a performance, as this will give us an idea of the extent of the deviation from the mean that we may expect when the standard is universally adopted. Thus, at a performance of Beethoven's first pianoforte concerto the piano was at 440, to which the orchestra tuned within 1 cycle/sec.; when accompanying the soloist the orchestra accommodated itself to his pitch, but rose about 2 cycles/sec. higher in the tutti. On the other hand, a rather regrettable state of affairs was disclosed in a broadcast of Bach's oboe concerto. Oboe and orchestra tuned to 440, but when the piano came in it was at 435 ! It says much for the string players that they overcame this handicap by playing at 439 (to which the oboe had dropped) when accompanying the solo, and at 435 in the ensembles with piano.

The pitch generally tends to rise above the tuning-note during forte passages, probably because the notes of the wind instruments tend to rise in pitch with the blowing pressure. The tests also showed—what has been long known—that the human voice at its best is by no means so precise in maintaining constancy of pitch as the mechanical instruments.

Madella (1939) has also recorded the extent of the deviation from the tuning pitch during a broadcast performance. Whereas the piano varies by one half-cycle per sec. and is able to hold the orchestra to that limit in concerti, a symphony orchestra (playing without the piano) usually fluctuates from the tuning pitch by 2 or 3 cycles/sec., though there was no general upward tendency. Evidently the wind had taken care to warm itself up before the performance !

The change of pitch of wind instruments with atmospheric conditions is a thing which has always militated against the adoption of universal pitch. Whereas the temperature effect on stringed instruments is small, the frequency of a column of air in vibration increases noticeably as the temperature of the contained air rises. While this rise can be readily predicted when the air is all at one temperature, the matter is complicated for orchestral wind instruments by the fact that the aerial contents are set in vibration by the player's breath. The temperature of the breath in expiration being more or less constant, the change of pitch with the temperature of the *atmospheric* air is much less in such cases, and depends on a number of other uncertain factors, such as the extent to which

the outside air can penetrate the instrument and the loss of heat by conduction through and radiation from the walls of the tube in which the air is confined. The oboe should be less affected on the average than the brass by atmospheric changes, since the former is smaller and enclosed in thick wooden walls instead of thin metal. Moreover, on the wood-wind the notes involving a short speaking-length would remain more constant than those for which the player closes most of the side-holes. It is, of course, with the object of attaining steady conditions that the wind players blow through their instruments for a minute or two before tuning. On the clarinet, where transposition demands the alternative use of two or more instruments, the sudden change from a "hot" to a "cold" instrument is particularly trying to a player who wants to keep the pitch up.

As far as organs and orchestras playing indoors are concerned, the temperature troubles are likely to be overcome when the adoption of air-conditioning becomes general and the temperature of concert halls is not allowed to vary more than a degree or two from 65° F. I should explain, for the benefit of those unacquainted with the mechanics of orchestral instruments, that whereas the strings and the trombone are "free-note" instruments in the sense that inaccuracies in the initial tuning can be offset by the player himself for each note played (if he plays by ear), the other wind instruments of the modern orchestra are "fixed-note" instruments, meaning that each length of the air column corresponds to a definite note, which can only be varied (apart from slight changes caused by pressure on the reed, if any) by altering the length of the column of air, which is accomplished by pushing the joints in more or less (on the same principle as those of a fishing-rod). Such an adjustment has to be made at the beginning of a performance and is strictly valid for one note only, since the change in question is *additive* and not *factorial*, as it would need to be if the whole instrument were to remain in just intonation.

§4. ELECTROPHONIC INSTRUMENTS

An account of the development of musical instruments in which the notes are produced synthetically was given in the last *Report*. To this we have only to add one 1939 paper on a violin with electrostatic pick-up from the strings and delivering the amplified sound into a loud-speaker, devised by Stauffer.

REFERENCES

BRITISH STANDARDS INSTITUTION, 1939. *British Standard Concert Pitch*, no. 880.
KAYE, G. W. C., 1938. *Nature, Lond.*, **142**, 820.
KAYE, G. W. C., 1939. *Nature, Lond.*, **143**, 905.
MADELLA, G. B., 1939. *Alta Frequenza*, **8**, 300.
MEINEL, H., 1939. *Akust. Z.* **4**, 89.
POL, B. VAN DER, and ADDINK, C. C. J., 1939. *Philips' Tech. Rev.* **4**, 205.
RICHARDSON, E. G., 1940. *J. Roy. Soc. Arts*, **88**.
SAUNDERS, F. A., 1940. *J. Franklin Inst.* **229**, 1.
STAUFFER, L. H., 1939. *J. Appl. Phys.* **10**, 96.
WILLIAMSON, C., 1939. *J. Acoust. Soc. Amer.* **11**, 216.
YOUNG, H. W., 1939. *J. Acoust. Soc. Amer.* **11**, 134.

PHOTO-ELECTRIC SPECTROPHOTOMETRY AND ITS USE IN INDUSTRY

By W. D. WRIGHT, A.R.C.S., D.Sc.,
Imperial College of Science and Technology

§1. INTRODUCTION

THIS article was to have been written by Professor A. C. Hardy of the Massachusetts Institute of Technology. Professor Hardy was the originator of the photo-electric spectrophotometer that is now being manufactured in the United States by the General Electric Company, and it was felt that in inviting him to contribute an article on this subject an opportunity would be provided of demonstrating the very great potentialities of a rapid method of spectrophotometry, a subject which has so far received little attention in this country. For various reasons Professor Hardy was not in a position to prepare a suitable article in the time available, but he kindly supplied some notes. The present writer has prepared the following short *Report*, partly based on these notes. Although it has the serious disadvantage of being largely second-hand, it may still serve to stimulate some interest in the subject among industrialists here. Acknowledgement must be made to Professor Hardy and to the other authors of a series of papers published in the *Journal of the Optical Society of America* in October 1938 for most of the information given below.

§2. THE PROBLEM OF RAPID SPECTROPHOTOMETRY

A spectrophotometer aims at measuring the energy distribution through the visible spectrum in a beam of light which may originate from a self-luminous source or may have become coloured by diffuse reflection from an opaque surface or by transmission through some solid or liquid colour filter. The problems with which we are chiefly concerned arise most frequently in the case of diffuse reflection, and to simplify the discussion we will confine ourselves to that case. The essential problem then is: being given a coloured surface, to determine the amount of light reflected from it at each point through the visible spectrum, relative to what would be reflected from a perfectly white diffusing surface. For the latter, a surface of magnesium oxide is generally considered to be a satisfactory standard, although for the highest accuracy special precautions in its preparation are called for (Gibson and Keegan, 1938).

To obtain the required data, a series of photometric matches is required between monochromatic beams; these may be made either visually or by means

of a photo-electric cell, but in neither case will the spectral sensitivity curve of the eye of the observer or of the photo-cell affect the result, always assuming that the absolute sensitivity is adequate throughout the spectrum. The fact that no standard eye or photo-cell is needed is an important advantage in spectrophotometry.

Any system of measurement that is to have a wide application must be capable of recording data for dark surfaces as well as light. The first and most serious difficulty encountered is therefore the very small amount of light sometimes available, so much so that visual observations are almost automatically ruled out as being, in these cases, quite impossible. The second difficulty is the time required for a complete run through the spectrum. Visual observations are certain to be long, tedious and fatiguing, and even photo-electric records in which the energy has to be measured point by point through the spectrum are likely to be lengthy. The importance of the time factor cannot be over-emphasized. If spectrophotometric curves are to be put to their full use, as, for example, in the dyeing industry, it is probable that very many curves would be required each day. An hour or more per curve is therefore out of the question, and to fulfil these needs adequately a self-recording instrument is almost essential.

§ 3. THE GENERAL ELECTRIC RECORDING SPECTROPHOTOMETER

It is not proposed to discuss the details of this instrument here; for these, reference may be made to Hardy (1935, 1938) and Michaelson (1938). For the purpose of this article it need only be stated that the optical system consists of two monochromators in series, the second serving to filter out stray light, and beyond the final slit there is a polarizing photometer unit. The light emerging from the system falls alternately on the sample under test and on the standard white surface as a Wollaston prism is rotated, and the relative intensities of the light falling on the two specimens is governed by the rotation of the analysing prism over a total range of 45°, this prism being mounted inside the hollow shaft of a synchronous motor. The light reflected from the sample and the standard is picked up by an integrating sphere and thence directed into a photo-cell unit placed behind a window in the sphere. The A.C. current from the cell is amplified and applied to the synchronous motor, the circuits being designed so that the motor maintains the analysing prism in the position necessary to keep a continual state of equality between the light reflected from the sample and from the standard. It is then only necessary to arrange a mechanism to vary the wave-length continuously and to attach a recording drum and recording pen to this mechanism and to the photometer unit for a spectrophotometric curve to be quickly and automatically traced out.

It is obvious that the successful development of an instrument of this complexity calls for extremely careful attention to the design of the various parts and requires a very high degree of workmanship, especially of the mechanical parts of the instrument. This will be the more readily appreciated when it is realised

that the significant current from the photo-cell may, under certain adverse circumstances, only be of the order of 10^{-15} amp. In spite of this, a successful apparatus has been produced, although skilled handling and check calibrations are required for results of the highest accuracy.

§ 4. THE OPERATION OF THE SPECTROPHOTOMETER

For certain applications it may be quite legitimate to sacrifice accuracy for speed, but unless an instrument were capable, when desired, of giving an accurate spectrophotometric curve, many of its advantages would be lost. The precautions that may then be necessary are described by Gibson and Keegan (1938) in a report on the operation of the General Electric instrument at the Bureau of Standards. These precautions include records, on the same chart as that of the curve taken for the sample under test, of the 100 per cent reflection curve, the zero-intensity curve and a curve showing the absorption of didymium glass. This last, because of its sharp absorption maxima, is used to provide a check of the wave-length calibration. Where corrections are necessary to give a true spectrophotometric curve, the extra time required for them must be taken into consideration when computing the total time occupied in recording the curve. In addition to checks on the calibration of the instrument, constant attention to the adjustment of the various units is needed to ensure reproducibility and the highest accuracy.

It is apparent that the operation of a recording spectrophotometer is a highly skilled job requiring the attention of an expert. Yet it would be a mistake to exaggerate the difficulties. As an example in which many of the more awkward corrections can be omitted, the case may be quoted of the comparison of two very nearly identical samples. In this problem, which is of frequent occurrence, small errors of adjustment and calibration only produce errors of the second order in the differences between the two curves.

One handicap of the G.E. spectrophotometer as described in 1938 is the difficulty of separating out the specular component from the diffuse component of the light reflected by an opaque test sample. This difficulty is inherent in the use of an integrating sphere to collect the reflected light, and since this is necessitated by the small amount of light sometimes available, any method of investigation which picks up a more restricted cone of light is unlikely to be of general application to both light and dark samples.

§ 5. APPLICATIONS OF SPECTROPHOTOMETRIC DATA

There are three main uses to which spectrophotometric data may be put, namely: (*a*) as an indication of the chemical nature of the dye or pigment that colours the medium, (*b*) as the basis from which the colour itself can be calculated in terms of the standard C.I.E. system, and, most valuable of all, (*c*) as the basic data from which to compute the colour produced by a mixture of two or more dyes or pigments.

Each industry has its own characteristic problems to solve, but their general type may be illustrated from those described by Foss (1938) in connection with the printing-ink industry.

Although the colouring properties of a pigment depend in the first place on the absorption spectrum of the pigment itself, it is not sufficient to measure this alone. It is obvious that the colour of the surface on which the ink is applied must also be taken into account, and one method of doing this might be to measure the spectral reflection curve of the surface. But this again is not sufficient, since other physical characteristics play a part, notably the extent to which the ink penetrates into the support, and the manner in which the gloss is affected by the application of the ink.

With such a variety of factors at work, it is not surprising that the safest road to a satisfactory result appears to be to take advantage of the technique acquired by the experienced craftsman. The scientific approach requires the investigation of each factor independently and then the assessment of the effect of their simultaneous operation. This may prove an involved and tedious process, but it must in the end be the most successful. As an instance of the systematic examination of such problems, the spectrophotometric curves have been measured for inks applied both to white opal glass and to black glass. These measurements enable the transparent and opaque components of the ink film to be separated, in each case no penetration of the ink film having taken place. When such data have been recorded for a series of thicknesses of the ink film and for a variety of inks, a great mass of information may be filed for future reference that should prove of inestimable value to manufacturers.

Spectrophotometric data allow a more thorough investigation of the opacity of an ink; they have proved valuable in indicating possible uses of new or previously unused pigments that would not otherwise have been apparent, and they may also serve to improve the materials themselves by showing up small differences between pigments made by varied processes. The photo-chemical reactions of bleaching and fading can be effectively studied with the spectrophotometer.

The most fundamental use to which the curves can be put is to evaluate the effect of mixing inks or dyes. In the case of inks, for example, this has been exemplified in an analysis of the best processing inks to be used in the various methods of colour reproduction. Spectrophotometric data provide the only means by which the colouring action of different inks can be determined and more exact colour reproduction secured.

The extent of the activity in this field in the United States cannot be estimated by the number of papers published on the subject. Most of the work is fundamental to the industries involved, so that many of the results are confidential and have not been published. Nevertheless, some idea may be obtained from the number of sheets of special recording paper used with the G.E. instrument; a conservative estimate indicates that at the present rate of development several million curves will have been recorded during the present decade. Yet, so far

as is known, the first curve on a self-recording instrument has still to be recorded in this country.

REFERENCES

Foss, C. E., 1938. *J. Opt. Soc. Amer.* **28**, 386.
Gibson, K. S. and Keegan, H. J., 1938. *J. Opt. Soc. Amer.* **28**, 372.
Hardy, A. C., 1935. *J. Opt. Soc. Amer.* **25**, 305.
Hardy, A. C., 1938. *J. Opt. Soc. Amer.* **28**, 360.
Michaelson, J. L., 1938. *J. Opt. Soc. Amer.* **28**, 365.

THE INFRA-RED SPECTRA OF POLYATOMIC MOLECULES

By HARALD H. NIELSEN,
Mendenhall Laboratory of Physics, Ohio State University, Columbus

§ 1. RÉSUMÉ OF EARLIER WORK

Nearly a quarter of a century has elapsed since high-dispersion measurements were first made on the infra-red absorption spectra of molecules. Since then the spectra of a great many diatomic and polyatomic molecules have been investigated, with the result that in the majority of cases complete or partial resolution of the bands has been attained. These spectra originate when by the absorption of light energy a molecule is raised from one rotation-vibration to another of higher energy. According to the quantum theory the frequency in wave numbers of the light absorbed will be given by the Bohr frequency relation

$$\tilde{\nu} = (E' - E'')/hc, \qquad \ldots\ldots(1)$$

E' and E'' being the upper and lower energy states, h Planck's universal constant $6{\cdot}624 \times 10^{-27}$ erg./sec. and c the velocity of light. Whenever it is possible to analyse the absorption bands in a spectrum, i.e., to identify the lines of the bands with definite quantum transitions, it is possible to determine the rotation-vibration energy-level pattern of the molecule in which the spectrum originates.

Theoretically these energies are the eigenvalues of the Schrödinger equation

$$(H - E)\Psi = 0, \qquad \ldots\ldots(2)$$

in which H is the quantum-mechanical Hamiltonian operator, which may be derived with the aid of the classical Hamiltonian function for the molecule. In considering the rotation-vibration energies of a molecule it may be supposed that the atomic nuclei assume a stable configuration under the influence of the force-field of the surrounding electrons. To the approximation in which we are interested it will then be legitimate to neglect the presence of the electrons in the molecule entirely and consider only the motion in space of the atomic nuclei. The classical Hamiltonian will therefore consist only of three parts: the kinetic energy of translation of the molecule as a whole; the kinetic energy of the atomic nuclei oscillating about their positions of equilibrium in a set of body-fixed coordinates free to rotate relative to a set of space-fixed axes; and the potential energy of the nuclei in the force-field of the electrons.

The actual potential energy is seldom known for a polyatomic molecule, since it can only be deduced when the problem of the electronic energies of the molecules has been solved. As an approximation, potential energy relations which are

perfectly general functions of the coordinates and which contain constants to be determined so as to give the correct positions of the fundamental vibration frequencies are sometimes used. Such functions frequently have the disadvantage that they contain more independent constants than can be determined from the experimental data. Simplifying assumptions which will supply additional relations between these constants must therefore be resorted to and the actual form of the potential energy relation for a molecule, as well as the number of independent constants it will contain, must depend upon the exact nature of the assumptions which are adopted. One such assumption which has justified itself in many instances is that of valence forces. This hypothesis supposes the forces between the atomic nuclei to be directed along the valence bonds and between the valence arms. Another simplification which in general has been less successful is that of central forces. In the case of the polyatomic hydrides (CH_3Cl, NH_3 etc.) one or more of the hydrogen atoms may be replaced by the deuterium isotope. Such isotopic molecules have the same force constants, but because of the altered reduced masses the vibration frequencies will in general be quite different. By determining the frequencies of oscillation of several isotopic molecules, a sufficient number of relations can be obtained so that, in principle at least, all the constants which occur in a potential energy relation which is perfectly general in the coordinates can be determined.

The most convenient method of obtaining the correct form of equation (2), once the classical Hamiltonian has been determined, is probably that described by Wilson and Howard (1936),* although it may also be arrived at by using the variational method of Schrödinger. It is convenient always to write the quantum-mechanical Hamiltonian H in the form $H_0+\lambda H_1+\lambda^2 H_2$, where H_0, λH_1 and $\lambda^2 H_2$ are the zeroth, first and second order terms. In an actual case H_0, λH_1 and $\lambda^2 H_2$ would, respectively, contain the terms

$$H_0=(h\nu/2)\{(8\pi^2u^2/h^2)[(C_0/A_0)P_x^2+(C_0/B_0)P_y^2+P_z^2] - (1/\nu)\Sigma_j\nu_j(\partial^2/\partial\xi_j^2-\xi_j^2)\}; \qquad \ldots\ldots(3\,a)\dagger$$

$$\lambda H_1=h\nu\{(2\pi^2u^3/h^2)(C_0/A_0)^{3/2}(\nu/\nu_j)^{\frac{1}{2}}\xi_jP_x^2+\ldots-(4\pi^2u^4/h^2)(C_0/A_0)p_xP_x + \ldots-(u^2\alpha_j/2)\partial/\partial\xi_j+\ldots+u(c_j\xi_j^3+\ldots)\}; \qquad \ldots\ldots(3\,b)$$

$$\lambda^2H_2=h\nu\{(4\pi^2u^4/h^2)(C_0/A_0)^2(\nu/\nu_j)\xi_j^2P_x^2+\ldots-(4\pi^2au^3/h^2)(C_0/A_0)^{3/2} \times(\nu/\nu_j)^{\frac{1}{2}}\xi_jp_xP_x+bu^2(\nu/\nu_j)^{\frac{1}{2}}\xi_jP_x+\ldots+p_x^2/2A_0 +\ldots+d(\partial/\partial q_j\,.\,q_j+q_j\partial/\partial q_j)+\ldots+g_j\xi_j^4+\ldots \qquad \ldots\ldots(3\,c)$$

* In a recent paper Darling and Dennison (1940) have pointed out that the equation of Wilson and Howard is not Hermitian and that an equation of a slightly different form should be used. A derivation of this equation has been promised in the immediate future by Darling. It appears, however, that to second order of magnitude there will be no significant difference between the two equations.

† The energy of translation of the molecule as a whole is also a part of H_0. It is non-periodic and contributes nothing to the vibration-rotation energies and may always be separated off from the remaining part of H_0.

In the above relations A_0, B_0 and C_0 are the equilibrium values of the principal moments of inertia of the molecule, α_j, c_j and d_j are constants, u^2 is a parameter of smallness equal to $(h/4\pi^2\nu C_0)$, P_x, P_y and P_z are the angular momentum operators and p_x, p_y and p_z are the internal angular momentum operators consisting of such terms as

$$[(\nu_k/\nu_j)^{\frac{1}{2}}\xi_j\partial/\partial\xi_k - (\nu_j/\nu_k)^{\frac{1}{2}}\xi_k\partial/\partial\xi_j],$$

where the ξ_k are the normal coordinates. The first term in equation (3 *a*) contains u^2 as a factor, and might therefore appear small compared to the second term. The operators P_x^2, P_y^2 and P_z^2, however, have eigenvalues quadratic in the rotational quantum numbers, so that the first term may be regarded as of the same order of magnitude as the second.

Equation (2) in its entirety never lends itself to an exact treatment since it is not separable in the coordinates. When, however, the higher order terms are neglected, and H is replaced by H_0, the equation (2) becomes separable and an exact solution may nearly always be obtained. The contribution to the energy arising from the presence of λH_1 and $\lambda^2 H_2$ may then be approximated by the methods of the perturbation theory.

The eigenvalues, E_0, of equation (2), where H_0 is used instead of H, are given for a definite value of the vibration quantum numbers V and the quantum number J by the roots of the secular determinant $|(K|\,H_0|\,K') - E_0\delta_{KK'}| = 0$, δ_{KK} being the Kronecker symbol, which has the elements

$$(V,J,K,M|\,H_0|\,V,J,K,M) = \sum_{V_i=1}^{s}(V_i+\tfrac{1}{2})h\nu_i + [J(J+1)-K^2] \times (h^2/16\pi^2)(1/A_0+1/B_0) - K^2h^2/8\pi^2C_0; \qquad \ldots\ldots(4\,a)$$

$$(V,J,K,M|\,H_0|\,V,J,K\pm 2,M) = (h/32\pi^2)\{[J(J+1)-K(K\pm 1)] \times [J(J+1)-(K\pm 1)(K\pm 2)]\}^{\frac{1}{2}}(1/A_0 - 1/B_0). \qquad \ldots\ldots(4\,b)$$

In the above, the quantum numbers V and J take only integral values; K also is an integer, taking all values from $-J$ to $+J$.

In one group of molecules to which a great many interesting examples belong, the moments of inertia A_0 and B_0 are the same. In such cases the non-diagonal elements (4 *b*) of the above secular determinant vanish, so that the energies are exactly the $(J+1)$ different diagonal elements (4 *a*). In molecules of this type the nuclear oscillations are essentially of two types: those where the electric moment oscillates along the axis of symmetry, known as parallel-type bands, and those where the electric moment varies perpendicular to the axis of symmetry, known as perpendicular-type bands. In the former case the selection rules for transitions from one rotation-vibration state to another allow ΔJ to be ± 1, 0 and $\Delta K = 0$, and for the latter case the selection rules are known to be $\Delta J = \pm 1$, 0 and $\Delta K = \pm 1$. A parallel-type band consists in general of a rather strong central line surrounded on either side by a set of uniformly spaced lines, the separation between any two being on the average $h/4\pi^2A_0c$. The structure of such a band remains essentially the same for all polyatomic molecules, the only real difference being in the intensity

with which the central line will occur. It is easy to show that as the moment of inertia C_0 approaches zero (viz., linear molecules) this line approaches zero in intensity. In figure 1 is shown the appearance of such a parallel-type band for a hypothetical molecule with moments of inertia $A_0 = B_0 = 5 \times 10^{-40}$ gm.-cm^2 and $C_0 = 1{\cdot}25 \times 10^{-40}$ gm.-cm^2. The nature of the perpendicular band has been studied by Dennison (1931) and by Bennett and Meyer (1928), and is known to consist of a group of parallel-type bands superimposed upon each other, the central branches being separated, to this approximation, by a frequency interval $(h/4\pi^2 c)(1/A_0 - 1/C_0)$. In figure 2 is depicted diagrammatically the structure of perpendicular bands. Five cases have been illustrated, beginning with the limiting case $A_0 = B_0 = C_0/2$, passing through the limiting case $A_0 = B_0 = C_0$ and ending with the limiting case $A_0 = B_0$, $C_0 = 0$. Between any two limiting cases is inserted an intermediate case to illustrate how the transition from one to the other takes place. The values of $A_0 = B_0$ are in each case taken to be 5×10^{-40} gm.-cm^2.

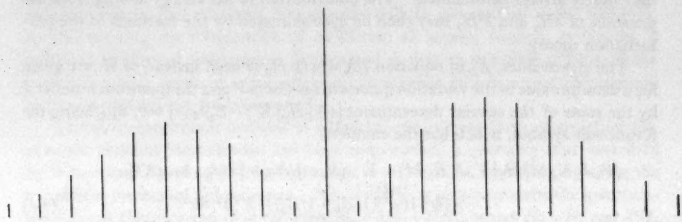

Figure 1.

In the more general type of molecule where the three moments of inertia A_0, B_0 and C_0 are all different, the off-diagonal matrix elements of the energy (4 *b*) will not vanish, so that the secular determinant $|(K|H_0|K') - E_0\delta_{KK'}|$ will have $2J+1$ different roots. This is equivalent to saying that the degeneracy in the quantum number K is here removed so that there will be $2J+1$ different energy values. Each one of these is customarily classified by use of an index number τ which runs from the value $+J$ to the value $-J$, so chosen that the largest energy value is identified with $\tau = +J$ and the lowest energy value with the index $\tau = -J$. The energy-level patterns for such molecules are in general complicated, although from the classical motion of an asymmetric top it is possible to draw certain conclusions concerning their appearance. For example, it may be demonstrated that rotation about the axis of the smallest moment of inertia and rotation about the axis of the largest moment of inertia represent stable types of motion for an asymmetric top, while rotation about the axis of the intermediate moment of inertia is an unstable type of motion. The analogy carries over into the quantum mechanics to the extent that for a given value of J those energies

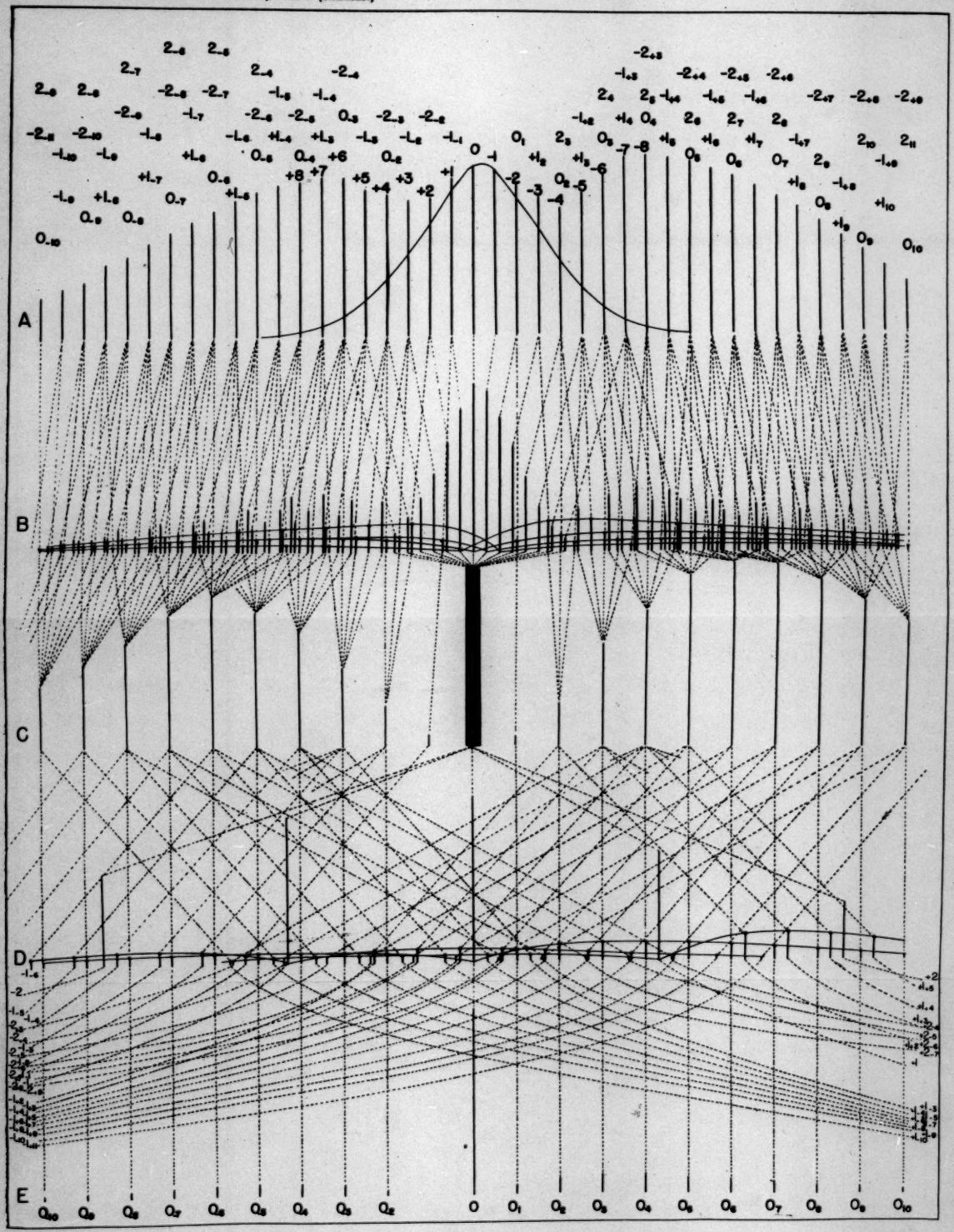

Figure 2.

To face page 44

where τ is approximately equal to J are very nearly expressed by the energy relations for a symmetric top spinning principally about its axis of smallest moment of inertia (i.e., $K=J$), and those energies for which τ is nearly equal to $-J$ are given approximately by the energy relations for a symmetric top spinning about its axis of largest moment of inertia (i.e., $K=J$). The levels identified with values of τ intermediate between $+J$ and $-J$ will, however, be badly approximated by either limiting case and correspond in a rough manner to rotation about the third

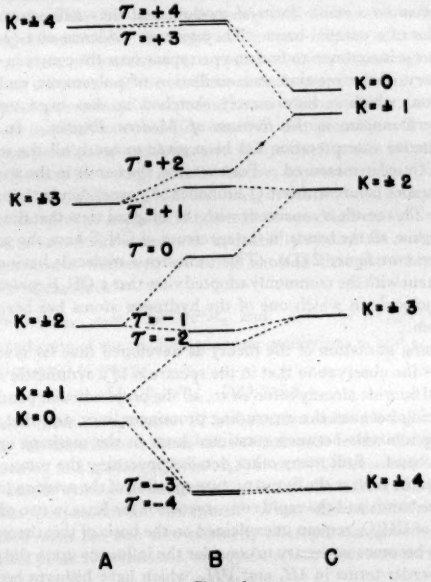

Figure 3.

axis. In figure 3 is shown a set of energy values for a molecule like the water-vapour molecule for the value $J=4$, and on either side are shown the energy levels of the limiting symmetric molecules.

The types of bands which may occur can no longer be spoken of as parallel bands or perpendicular bands except in a very limited way. This may be seen as follows. A molecular oscillation which is along the axis of the smallest moment of inertia will have the rotational characteristics of a parallel-type band for those states where the molecule is spinning principally about that axis, i.e., for states where

$\tau = +J$, but for the states where the molecule spins principally about the axis of the largest moment of inertia it will have the characteristics of a perpendicular band. A similar kind of reasoning applies to those oscillations of a molecule which are along the axis of the largest moment of inertia. These two kinds of oscillations will, in general, have quite different rotational characteristics, but may have similarities in certain respects. For example, they will both have Q-branches (which are often quite diffuse) at the centre. Oscillations which are parallel to the axis of the intermediate moment of inertia will be normal to both of the axes about which rotation represents a stable form of motion, and they will possess none of the characteristics of a parallel band. They give no evidence of a Q-branch at the centre; in fact it is common to find an open space near the centre in such bands.

The theory of the rotation and oscillation of polyatomic molecules to this approximation, which is here merely sketched in, has been very adequately discussed by Dennison in the *Reviews of Modern Physics*. In terms of it a quasi-quantitative interpretation has been given to nearly all the infra-red molecular spectra thus far measured. For example, the bands in the spectrum of CO_2 are either parallel bands without Q-branches or perpendicular bands like that in figure 2 (C). This result is consistent with the adopted view that the CO_2 molecule is linear. Again, all the bands in the spectrum of CH_3F have the general appearance of figure 1 or figure 2 (D). This indicates a molecule having $A_0 = B_0 \gg C_0$ and is consistent with the commonly adopted view that a CH_3F molecule is simply a methane molecule in which one of the hydrogen atoms has been replaced by a fluorine atom.

A distressing limitation of the theory as developed thus far is its inadequacy to account for the observation that in the spectrum of a symmetric molecule, like that of methyl fluoride already referred to, all the perpendicular bands do not have the same spacing between the protruding prominent lines, and that, for example, the frequency intervals between rotational lines in the methane spectrum vary from band to band. Still many other details concerning the rotational structure of infra-red bands, such as the fine-structure splitting of the rotation lines in certain of the methane bands and the rapid convergence of the lines in two of the bands in the spectrum of CH_2O, remain unexplained on the basis of this theory. For their clarification it becomes necessary to consider the influence upon the energy levels of the higher-order terms in λH_1 and $\lambda^2 H_2$, which have hitherto been neglected. The study of the infra-red spectra of polyatomic molecules thus enters upon a second stage in its development.

§ 2. THE PROBLEM OF THE INTERACTIONS BETWEEN ROTATION AND OSCILLATION

The influence of such terms as are present in λH_1 and $\lambda^2 H_2$ on the energy-level pattern of a diatomic molecule was studied by Fues as early as 1926 and their effect was found, in general, to be small. For a polyatomic molecule, the terms contained in λH_1 and $\lambda^2 H_2$ are very similar to those for a diatomic molecule, namely, higher order (anharmonic) terms in the potential energy, centrifugal

stretching terms and a set of terms (of which there are none analogous in the diatomic case) known as Coriolis interaction terms. The latter have the form $(p_x P_x/A_0)+(p_y P_y/B_0)+(p_z P_z/C_0)$. In a molecule where all the vibration frequencies are distinct and different, the corrections to the energy due to λH_1 and $\lambda^2 H_2$ are again found to be relatively small, their contribution being only in second approximation. In such cases Wilson and Howard have pointed out that the interactions may be fairly well taken into account in the energies by assigning to the rotating molecules moments of inertia which are effective only in a given state, so that theoretically the term values may still be arrived at by solving a secular determinant essentially the same as the one already referred to in § 1. The effective moments of inertia occur as parameters in the energies, and since they depend upon the nature of the oscillations and upon the extent of the excitation of the molecule, they may be expected to vary from state to state. In cases where degenerate oscillations occur (for example the optically active oscillations in CH_4 or the perpendicular oscillations in NH_3) the Coriolis interaction term, $\boldsymbol{p}.\boldsymbol{P}$, may contribute to the energy even in the first approximation. This was first pointed out by Teller and Tisza (1932) and later, in a more satisfactory manner, by Teller (1934) in an effort to explain the anomalous spacing between rotation lines in the bands of tetrahedrally symmetric and axially symmetric molecules. For methane-like molecules the work of Teller showed that the interaction term $\boldsymbol{p}.\boldsymbol{P}/A_0$ had the eigenvalues

$$\epsilon_+ = J\zeta_i h^2/4\pi^2 A_0, \quad \epsilon_0 = -\zeta_i h^2/4\pi^2 A_0 \quad \text{and} \quad \epsilon_- = -(J+1)\zeta_i h^2/4\pi^2 A_0$$

in the first excited state of the triply degenerate oscillations ν_3 and ν_4, so that each rotation-vibration level of energy

$$5h\nu_i/2 + J(J+1)h^2/8\pi^2 A_0$$

will be split into three components. The selection rules are $\Delta J = +1,\ 0,\ -1$, respectively, as the transition is from the normal state to the component states ϵ_+, ϵ_0 and ϵ_-, leading to a spacing $(\zeta_i h/4\pi^2 A_0 c)$ in cm.$^{-1}$ between two rotation lines. In the case of axially symmetric molecules, the Coriolis interaction term $p_z P_z/C_0$ affects only the perpendicular vibrations. When the normal coordinates ξ_i and ξ_j associated with the two dimensionally isotropic oscillations are replaced by the cylindrical polar coordinates ρ_i and χ_i the above term may be written

$$(\zeta_i h^2/4\pi^2 C_0)\partial^2/\partial\phi\,\partial\chi.$$

For the first excited states this has the eigenvalues

$$\epsilon_\pm = \pm\zeta_i(Kh^2/4\pi^2 C_0),$$

and the new selection rules, this time for K (the selection rule for J remains unchanged), state that $\Delta K = +1$ or -1, depending upon whether the transition is from the normal state to the component levels ϵ_+ or ϵ_-. The frequency interval between two Q-branches in a perpendicular band will then evidently take the modified value

$$(h/4\pi^2 c)\,[(1-\zeta_i)/C_0 - 1/A_0].$$

The quantities ζ_i occurring in the above relations are constants which depend upon the nature of the normal oscillation and will therefore, in general, vary from frequency to frequency. In order, therefore, correctly to evaluate the moments of inertia from a given band in the spectrum it is necessary to know the corresponding value of ζ_i. This quantity is frequently very complicated to evaluate. Fortunately, Teller has also demonstrated that $\Sigma\zeta_i$ for a given symmetry class is a simple quantity independent of the potential-energy constants and depending only upon the internuclear distances within the molecule and upon the masses of the atoms themselves. In a more rigorous application of this theory to certain polyatomic molecules Dennison and Johnston (1935) have evaluated these sums, showing that for molecules of the type XY_4, $\Sigma\zeta_i = \frac{1}{2}$ and for molecules of the types XY_3 and ZXY_3, $\Sigma\zeta_i$ is respectively equal to $(C_0/2A_0) - 1$ and $C_0/2A_0$. From these values of $\Sigma\zeta_i$ it is readily shown that the spacings between two lines summed over the two optically active XY_4 fundamental bands will be

$$\Sigma\Delta\nu = 3h/8\pi^2 A_0 c,$$

and the frequency intervals between Q-branches summed over the fundamental perpendicular bands of the XY_3 and ZXY_3 types of molecules will be respectively

$$\Sigma\Delta\nu = (h/4\pi^2 c) \times (3/C_0 - 5/2A_0) \quad \text{and} \quad \Sigma\Delta\nu = (h/4\pi^2 c)(3/C_0 - 7/2A_0).$$

Dennison and Johnston have, further, made calculations for these molecules on the values of the ζ's in the case of individual bands. It is interesting to note that the value of ζ associated in each case with the high-frequency fundamental is always much smaller (in fact very often approaching zero) than the ζ's associated with fundamentals of smaller frequency. This observation may be understood in the following manner. The degenerate molecular oscillation which gives rise to this small value of ζ is in each case one where principally the hydrogen atom and the carbon atom (or nitrogen atom) participate in the motion. As long as the mass of the hydrogen atom may be considered small compared to that of the carbon (or nitrogen) atom, the motion may be regarded essentially as one where the former move along the valence bonds. The induced electric moment will therefore be along the valence bonds and will, as a consequence, rotate with the same speed as the framework of the molecule. An apparent exception to this rule is that of the NH_3 molecule. Barker (1939) has resolved and given an analysis of the frequency ν_2 for that molecule. He finds, however, that the value of ζ_2 as determined by experiment is equal to $+0{\cdot}6$ as compared to the value $-0{\cdot}004$ calculated by Dennison and Johnston.* When the hydrogen atoms are replaced by deuterium atoms the above generalization no longer holds nearly so well, indicating that here the nuclear motions are only approximately along the valence bonds. These authors have further extended the theory satisfactorily to include certain of the higher energy states of these molecules. In a later work by Shaffer, Nielsen and Thomas (1939) it has been demonstrated how the first-order Coriolis splitting of any rotation-vibration state in tetrahedrally symmetric molecules may be computed, using the general methods of vector addition of quantum mechanics.

* See note on p. 65.

The method outlined by them has been applied specifically to the harmonic states $2\nu_3$ and $2\nu_4$ and to the combination state $\nu_3+\nu_4$.

For their purpose $\boldsymbol{p}$ is written $\zeta_3\boldsymbol{J}_3+\zeta_4\boldsymbol{J}_4$, where $\boldsymbol{J}_3$ and $\boldsymbol{J}_4$ are the internal angular momenta associated with ν_3 and ν_4 respectively. From a consideration of the oscillations ν_3 and ν_4 as three-dimensionally isotropic oscillators in terms of spherical polar coordinates it is easy to show that J_3 may take the values $V_3, (V_3-2), \ldots 0$ or 1 (and J_4 the values $V_4, (V_4-2), \ldots.$), where V_3 and V_4 are the vibrational quantum numbers associated with ν_3 and ν_4 respectively. The total internal angular momentum, $\boldsymbol{J}_2$, is then $\boldsymbol{J}_3+\boldsymbol{J}_4$, which evidently may take the values

$$(J_3+J_4),\ (J_3+J_4-1), \ldots (J_3-J_4).$$

The total angular-momentum vector J is the sum of the vectors $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, $\boldsymbol{J}_1$ being the angular momentum of the rotation of the molecular framework, which may assume the values

$$(J+J_2),\ (J+J_2-1), \ldots\ (J-J_2).$$

Applied to the state $2\nu_2$, $V_3=2$, so that J_3 may take the values $J_3=2$ and $J_3=0$. The total internal angular-momentum vector $\boldsymbol{J}_2$ is thus equal to $\boldsymbol{J}_2=\boldsymbol{J}_3$. In the instance where $J_3=2$, it may be coupled to J_1 in five different ways (see figure 4, C). Corresponding to the five values of $J_1((J+2), (J+1), \ldots (J-2))$, the Coriolis operator yields the five eigenvalues

$$\epsilon_1=J\zeta_3h^2/2\pi^2C_0,\quad \epsilon_2=(J-2)\zeta_3h^2/4\pi^2C_0,\quad \epsilon_3=-3\zeta_3h^2/4\pi^2C_0,$$
$$\epsilon_4=-(J+3)\zeta_3h^2/4\pi^2C_0 \quad\text{and}\quad \epsilon_5=-(J+1)\zeta_3h^2/2\pi^2C_0.$$

The selection rules for J are in this case more complicated than for the fundamental band. They are $\Delta J=0, -1$; $\Delta J=0, \pm1$; $\Delta J=0, +1$, as the transition is from the normal state to the component levels ϵ_1; ϵ_2, ϵ_3, ϵ_4; and ϵ_5 respectively. The case $J_3=0$ gives the eigenvalue $\epsilon_0=0$, but from the normal state to this component no transition is permitted. For the state $\nu_3+\nu_4$ the situation is still more complicated. Here $J_3=J_4=1$, so that J_2 may be equal to 2, 1 or 0. These may be coupled to J_1 in 5, 3 and 1 different ways respectively (see figure 4), corresponding to the values of J_1:—$J+2, \ldots J-2$; $J+1, \ldots J-1$; and J. There will correspondingly be nine different eigenvalues, which are as follows:—

For $J_1=J+2$ and $J_1=J-2$,

$$\epsilon_1=J(\zeta_3+\zeta_4)h^2/4\pi^2C_0 \text{ and } \epsilon_2=-(J+1)(\zeta_3+\zeta_4)h^2/4\pi^2C_0 \text{ respectively;}$$

for $J_1=J+1$,

$$\epsilon_{3,4}=\tfrac{1}{2}\{(J-1)(\zeta_3+\zeta_4)\pm[(J+1)^2(\zeta_3-\zeta_4)^2+4\zeta_3\zeta_4]^{\frac{1}{2}}\}(h^2/4\pi^2C_0);$$

for $J_1=J-1$,

$$\epsilon_{5,6}=-\tfrac{1}{2}\{(J+2)(\zeta_3+\zeta_4)\pm[J^2(\zeta_3-\zeta_4)^2+4\zeta_3\zeta_4]^{\frac{1}{2}}\}(h^2/4\pi^2C_0);$$

and for $J_1=J$ the roots $\epsilon_{7,8,9}$ are the roots of the cubic equation

$$(\epsilon h^2/4\pi^2C_0)^3+2(\zeta_3+\zeta_4)(\epsilon h^2/4\pi^2C_0)^2+[3/4(\zeta_3+\zeta_4)^2$$
$$-1/4(4J^2+4J-1)(\zeta_3-\zeta_4)^2](\epsilon h^2/4\pi^2C)_0-J(J+1)(\zeta_3+\zeta_4)(\zeta_3-\zeta_4)^2=0.$$

Their method also enables one to obtain the stabilized wave functions for these

states, thus permitting the calculation of the selection rules and the intensities. It is found that the selection rules for J are $\Delta J = 0, -1$ for transitions from the normal state to the component ϵ_1, $\Delta J = 0, +1$ to the component ϵ_2 and to all the other components $\Delta J = \pm 1, 0$.

Even in an unsymmetrical molecule, if there happen to be two or three normal oscillations belonging to different symmetry classes which are accidentally degenerate, the Coriolis interaction terms may yield large contributions to the energy. The importance was originally pointed out by Wilson (1936) and subsequently by Jahn (1938 a) with reference to tetrahedrally symmetric molecules, but their eigenvalues were first evaluated in the case of the formaldehyde molecule

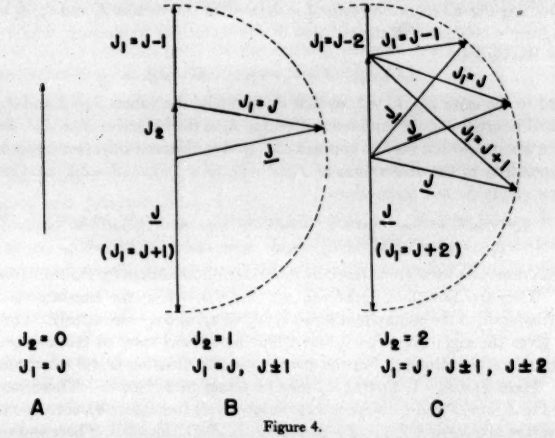

Figure 4.

in an effort to interpret two bands in the spectrum of that molecule. Here the frequency ν_6, which is in and out of the plane of the molecule, falls at 1165 $cm.^{-1}$ and the frequency ν_5, which is essentially a frequency of deformation of the O–C–H bond angles in the plane of the molecule, is found to be near 1278 $cm.^{-1}$ These frequencies may be regarded as perpendicular-type oscillations which are at right angles to each other and therefore evidently belong to different symmetry classes. It has been shown by Nielsen (1939) that these give rise to a strong Coriolis interaction which manifests itself in the spectrum as a rapid convergence of the Q-branches in ν_6 toward higher frequencies and in ν_5 toward lower frequencies. The positions of the lines in the two bands are given by the relation

$$\nu = \tfrac{1}{2}(\nu_5 + \nu_6) \pm (Kh/4\pi^2)(1/C_0 - 1/A_0) \pm [1/4(\nu_6 - \nu_5)^2 + (K\zeta h/4\pi^2 C_0)^2]^{\frac{1}{2}},$$

where

$$1/A_0 \approx \tfrac{1}{2}(1/A_0 + 1/B_0),\ 1/B_0 \text{ and } 1/C_0$$

are respectively the reciprocals of the largest, intermediate and smallest moments of inertia. The effect was verified by the measurements made on these frequencies by Ebers and Nielsen (1937). An even more interesting example is that of deutero-formaldehyde, where ν_4, the frequency where the D–C–D bond angle is deformed, falls at 1106 $cm.^{-1}$, becoming accidentally degenerate with the frequencies ν_5 and ν_6, which in this case fall at 990 $cm.^{-1}$ and 938 $cm.^{-1}$ respectively. The Coriolis interaction here produces an effect on these bands similar to the one produced in the ordinary formaldehyde bands, but at the same time it causes the parallel band ν_4 to become degraded toward the red. This is clearly indicated by the measurements by Ebers and Nielsen for that molecule on the shape of the Q-branch of ν_4 (Ebers and Nielsen, 1938). A similar type of Coriolis interaction occurs in tetrahedrally symmetric molecules between the doubly degenerate and optically inactive frequency ν_2 and the three-fold degenerate oscillation ν_4. In a series of papers by Childs and Jahn (1939) and by Jahn (1939) it has been demonstrated how this interaction may be made to account for the fine-structure splitting of the rotation lines in the band ν_4. The effect of the perturbation in the excited state is to destroy the identity between the three moments of intertia A_0, B_0 and C_0 to such an extent that the $(2J+1)$-fold degeneracy in the quantum number K is removed. The spectrum predicted theoretically by Childs and Jahn agrees reasonably well with the earlier measurements made by Nielsen and Nielsen (1935) on this frequency. In a recent paper Murphy (1940) has applied the development of Jahn also to the spectra of silane (SiH_4) and germane (GeH_4), with the result that something better than qualitative agreement between the experimentally observed bands of Steward and Nielsen (1935 a and 1935 b) and the predicted spectrum has been achieved. The most obvious result of the above work is that it accounts for the convergence toward higher frequencies of the lines in ν_4 in the spectrum of SiH_4 and for the convergence in the opposite sense for the lines in ν_4 in the GeH_4 spectrum by pointing out that in the former case ν_2 is a larger frequency than ν_4 and in the latter case the reverse is true. This conclusion is consistent with those already referred to concerning the formaldehyde molecules.

If the unperturbed wave functions of the states ν_4 and ν_2 be symbolically represented by $\psi(\nu_4)$ and $\psi(\nu_2)$ then the perturbed wave functions ψ_{I} and ψ_{II} will be linear combinations of the above two; for example: $\psi_{\mathrm{I}} = a_{\mathrm{I}}\psi(\nu_4) + b_{\mathrm{I}}\psi(\nu_2)$ and $\psi_{\mathrm{II}} = a_{\mathrm{II}}\psi(\nu_4) + b_{\mathrm{II}}\psi(\nu_2)$. When the frequencies ν_2 and ν_4 are widely separated, the ratios of a_{I} to b_{I} and a_{II} to b_{II} will be such that ψ_{I} is principally $\psi(\nu_4)$ and ψ_{II} is principally $\psi(\nu_2)$. In the case of SiH_4, the frequencies ν_2 and ν_4 are separated only by about 60 $cm.^{-1}$, so that the "mixing" of the wave functions is sufficient to induce an optical activity in the otherwise optically inactive frequency ν_2. Jahn (1938 b) has suggested this as the explanation of the second (higher frequency) maximum in the 10-μ absorption region in the spectrum of SiH_4. An estimate has been made by the writer of the ratio of the intensities of these two bands as predicted by theory, and the agreement with the experimental data

seems satisfactory. It would be of interest to make a search for the corresponding band ν_2 also in the spectrum of methane and germane. These bands would, of course, be much less intense than is the case in the SiH_4 spectrum.*

Coriolis interactions similar to those discussed in the foregoing paragraph may be expected in a great many other molecules. A case which at once suggests itself is that of water vapour, where the mutually perpendicular oscillations ν_1 and ν_3 fall near 3650 $cm.^{-1}$ and 3755 $cm.^{-1}$ respectively and are separated by only about 100 $cm.^{-1}$ It has, however, been shown by Darling and Dennison that the modulus of the internal angular momentum produced by these frequencies is very nearly equal to zero, so that the interaction will be very small. The frequencies ν_1 and ν_3 both arise from oscillations where the hydrogen atoms move along the O–H bonds. The conclusions of Darling and Dennison are therefore consistent with our earlier remark that in such cases the electric moment rotates with the same speed as the molecular framework.

From the foregoing considerations two convenient generalizations may be drawn concerning Coriolis interactions. The first of these is that whenever the atomic nuclei which participate in the oscillational motion may be regarded as moving along the valence bonds, as is essentially the case in X–H vibrations in polyatomic hydrides, the electric moment will rotate with the same speed as the molecular framework. This is equivalent to saying that in such cases the modulus of the internal angular momentum of the oscillation will be small and the Coriolis interaction will be at a minimum. The second generalization to be drawn is that two accidentally degenerate frequencies belonging to different symmetry classes may give rise to a strong Coriolis interaction between the rotational and the oscillational motion of the molecule, the effect of which is to cause rapid convergence in the higher frequency band toward lower frequencies and in the lesser frequency band toward higher frequencies. When, as in CD_2O, there are three accidentally degenerate oscillations each belonging to a different symmetry class, the rule still holds for the frequencies taken two and two.

Another type of interaction for which it is necessary to make corrections when analysing the spectrum of a polyatomic molecule is the one arising from the centrifugal stretching of the molecule. In recent papers by Randall, Dennison, Ginsberg and Weber (1937) and by Fuson, Randall and Dennison (1939) measurements have been reported on the rotation spectrum of water vapour and heavy-water vapour. It has virtually been possible for them to resolve the spectrum completely in this region, frequency intervals as small as 0·25 $cm.^{-1}$ having been recorded. The authors have been successful in identifying the measured lines with quantum transitions in most cases, and in this manner they have determined the positions of the rotation levels up to the value $J = 12$. It is found, however, that the levels are not satisfied by the energy relations for the asymmetric top with rigid moments of inertia, but that the moments of inertia must be regarded as effective only in a given rotation state. The corrections to the energy

* See note 2 on p. 65.

due to the centrifugal stretching of the molecule have been computed by these authors in the following semi-classical manner. For states where $\tau \approx +J$ the molecule, as we have seen, will be spinning about the axis of the smallest moment of inertia, which for a molecule like the water-vapour molecule will be parallel to the line joining the two hydrogen atoms in the non-rotating state. If the normal distance from the oxygen atom to the line joining the two hydrogen atoms be designated by y_0, the smallest moment of inertia, A_0, will be seen to be μy_0^2, where μ is the reduced mass $2mM/(2m+M)$, m and M being the masses of the hydrogen and oxygen atoms respectively. The energy of the rotating rigid molecule will, of course, be $(W_0/hc) = J^2h^2/8\pi^2A_0c$. Designating the change in the distance between the hydrogen atoms by q and the displacement of the oxygen atom relative to the centre of gravity perpendicular to the axis of A by y, it is easily shown that the rotational energy and that part of the potential energy effected by the centrifugal forces when the rotation is entirely about A will be

$$(W/hc) = J^2h^2/8\pi^2Ac + \tfrac{1}{2}(ay^2 + bq^2 + 2dqy), \qquad \ldots\ldots(5)$$

where a, b and c are potential-energy constants. Adopting the notation

$$\alpha = h/8\pi^2A_0c, \quad \beta = ((ab-d^2)/2b)y_0^2 \quad \text{and} \quad \eta = y/y_0,$$

one may, with the aid of the two Lagrange equations of motion, rewrite (5) in the form

$$W/hc = \alpha J^2/(1+\eta)^2 + \beta\eta^2$$

and arrive at the relation $\eta(1+\eta)^3 = \alpha J^2$, which determines η in the above equation for W/hc in terms of α, β and J. The correction to the energy will then be

$$\delta W/hc = (W - W_0)/hc = -\eta^2(1+3\eta+\eta^2),$$

which to first approximation will be seen to be $-\alpha^2J^4$. A similar development can be carried out for the energy levels where $\tau \approx -J$. While the corrections to the energy computed on this basis are always of the right order of magnitude, the model adopted by Randall, Dennison, Ginsburg and Weber is, of course, far too simple to lead to results which are quantitatively right.

The investigations thus far enumerated in § 2 have all represented efforts to account for anomalies of one kind or another observed in the spectra of polyatomic molecules which remain unexplained by the theory in zeroth approximation. The additional terms which are here introduced are all terms which occur in the first- or second-order Hamiltonian λH_1 and $\lambda^2 H_2$. In most cases they have been selected as the most important terms in λH_1 and $\lambda^2 H_2$, all others having been neglected, or they have been arrived at by use of a model much too naïve; they consequently make no pretence of representing complete theories of these molecules. In fact the foregoing work all serves to emphasize the extreme importance of developing theories for various polyatomic models, accurate to the same degree of approximation as that developed by Fues for diatomic molecules.

It has been assumed in § 1 that the quantum-mechanical Hamiltonian is already known to second order of magnitude and that the eigenvalues and eigenfunctions to H_0 have been determined. The corrections to the energy arising

out of the presence of H_1 and H_2 are most readily arrived at through the methods of the perturbation theory in quantum mechanics. As is well known, the first- and second-order energies computed in this manner are the following:

$$E_1=(V, R|H_1|V, R) \qquad \ldots\ldots(6\,a)$$

$$E_2=(V, R|H_2|V, R)+\sum_{V'R'}{}'(V, R|H_1|V', R')(V', R'|H_1|V, R)/(E_{V,R}-E_{V'R'}) \qquad \ldots\ldots(6\,b)$$

where V represents the vibration quantum numbers and R the rotational quantum numbers.

The evaluation of the relations (6 *a*) and (6 *b*) is a formidable undertaking when one considers the larger number of terms which in general may occur in λH_1 and $\lambda^2 H_2$. The only terms in λH_1, however, which can contribute in first approximation are the Coriolis terms $p_\alpha P_\alpha$ arising from any degenerate oscillations which are present. It is, moreover, always possible to find linear combinations of the eigenfunctions of H_0 such that the matrix of the above terms will have only diagonal elements. This suggests the effectiveness of transforming the Hamiltonian H by means of a contact transformation THT^{-1} into $H_0'+\lambda H_1'+\lambda^2 H_2'+\ldots$, where now H_1' will, to second approximation, contain only the Coriolis terms referred to above. The evaluation of the second-order energies will thus, in principle, be reduced to a first-order perturbation calculation, where the eigenfunctions to be used are the stabilized wave functions of H_0', i.e., the linear combinations of zeroth-order wave functions which diagonalize $H_0'+\lambda H_1'$. We shall represent the transformation function $e^{i\lambda S}$ and its inverse $e^{-i\lambda S}$ by $T(\lambda)$ and $T(\lambda)^{-1}$. When expanded to second order

$$T=1+i\lambda S-\tfrac{1}{2}\lambda^2 S^2-\ldots \qquad \ldots\ldots(7)$$

The transformed Hamiltonian will become to second order

$$THT^{-1}=H'=H_0'+\lambda H_1'+\lambda^2 H_2', \quad \text{where} \quad H=H_0+\lambda H_1+\lambda^2 H_2.$$

When the above expansion for T and its inverse are used in the indicated transformation one obtains, by equating like powers of λ, the following result:

$$\left.\begin{aligned} H_0' &= H_0, \\ H_1' &= H_1-i(H_0S-SH_0), \\ H_2' &= H_2+(i/2)[S(H_1+H_1')-(H_1+H_1')S]. \end{aligned}\right\} \qquad \ldots\ldots(8)$$

Those terms in H_1 which it is desired to remove will be functions of the normal coordinates q_k or the conjugate momenta p_k. Each term will be multiplied by a constant coefficient or a coefficient which is a function of the angular momentum operators P_α. When the coefficients are functions of the momenta they may still be treated as constants so long as at least two of the moments of inertia are alike. This becomes clear since the rotational part of H_0 is then a diagonal matrix so that H_0 commutes with P_α, i.e., $(H_0P_\alpha-P_\alpha H_0)=0$. In the case of an asymmetric molecule the transformation may still be carried out, but here some further caution is required.

In the case of symmetric molecules the basic transformation functions S_ρ

required to remove the type of terms occurring in λH_1, together with the corresponding values of $i(H_0 S_\rho - S_\rho H_0)$, have been evaluated by Shaffer, Nielsen and Thomas (1939) and are tabulated in table 1.

Table 1. Basic transformation functions S_ρ required to remove the types of terms occurring in $H^{(1)\prime}$ together with the corresponding values of $i(H^{(0)} S_\rho - S_\rho H^{(0)})$

ρ	S_ρ	$i(H^{(0)} S_\rho - S_\rho H^{(0)})$
1	$(-p_k/h\nu_k)$	q_s
2	$(q_k/h\nu_k)$	q_k
3	$(-2/h\nu_k)\,(p_k^2/3 + \frac{1}{2}q_k p_k q_k)$	q_k^3
4	$(1/h)\,(\nu_\sigma q_r q_s + \nu_r p_r p_s)/(\nu_r^2 - \nu_s^2)$ *	$q_r p_s$
5	$-\{(2\nu_r^2 - \nu_s^2) q_r^2 p_s + \nu_r \nu_s (q_r p_r + p_r q_r) q_s + 2\nu_r^2 p_r^2 p_s\}/h\nu_s(4\nu_r^2 - \nu_s^2)$	$q_r^2 q_s$
6	$-\{\nu_r(\nu_r^2 - \nu_s^2 - \nu_t^2) p_r q_s q_t + \nu_s(\nu_s^2 - \nu_r^2 - \nu_t^2) q_r p_s q_t - 2\nu_r \nu_s \nu_t p_r p_s p_t + \nu_t(\nu_t^2 - \nu_r^2 - \nu_s^2) q_r q_s p_t\} h^{-1}(\nu_r^4 + \nu_s^4 + \nu_t^4 - 2\nu_r^2\nu_s^2 - 2\nu_s^2\nu_t^2 - 2\nu_r^2\nu_t^2)^{-1}$	$q_r q_s q_t$
7	$(q_r q_s + p_r p_s)/h(\nu_r - \nu_s)$	$q_r p_s - p_r q_s$

* S_4 is the function which removes from $H^{(1)}$ the Coriolis interaction terms which are of the type $(q_r p_s - p_r q_s) P_\sigma$. When these terms originate with degenerate oscillations, the difference $\nu_r - \nu_s$ occurring in the denominator vanishes. Obviously such terms cannot be removed from $H^{(1)}$ by this method.

The complete S function will evidently be a sum of terms each of which is a basic S_ρ multiplied by an appropriate coefficient and will contain as many elements as there are terms in H_1 to be removed, each S_ρ being chosen to remove a single term in λH_1. The application of a contact transformation to this type of problem may, of course, be thought of as another method of carrying out perturbation calculations. The real advantage comes from the fact that nearly all the terms in λH_1 give no contribution in first approximation and that combinations of the zeroth-order wave functions may always be found which will diagonalize those terms in H_1 which do contribute in first order. When the transformation is not applied, but equation (6 *b*) is evaluated as it stands, it is almost necessary to make use of the fact that the rotational separations are small compared with the vibrational separation, so that $E_{V,R}$ may be replaced by E_V. This approximation becomes unnecessary when the transformation is carried out.

Transformations of the above kind have been used by Van Vleck (1929) in calculating σ-type doubling in diatomic molecules and by Jordahl (1934) in calculations on paramagnetism. Their application to the problem of the interactions between rotation and vibration in polyatomic molecules was suggested in the original papers by Wilson and Howard.

The first polyatomic molecular model on which calculations were ever made on the contributions to the energy due to such terms as occur in λH_1 and λH_2 was the linear XY_2 type studied first by Adel and Dennison (1933 a and 1933 b) and subsequently by Weinberg and Eckhart (1937). In our notation H_0 will contain the squares of the linear momenta, the quadratic terms in the potential

energy and the rotational energy, whilst λH_1 will contain the first-order centrifugal stretching terms which are linear in the normal coordinates, the first-order Coriolis terms and the cubic anharmonic terms in the potential energy, and finally, $\lambda^2 H_2$ will embrace the second-order centrifugal stretching terms which are quadratic in the normal coordinates, second-order Coriolis force terms and the quartic terms in the potential energy. In the case of linear molecules† the first-order energy corrections $E_1 = 0$. To simplify the evaluation of the second-order energies, particularly the second term in (6*b*), Adel and Dennison have introduced the following short cut with reference to the first-order or linear centrifugal stretching terms which can contribute only in second approximation, and then by an amount only of the same size as that contributed by the second-order centrifugal stretching terms. These terms effect a change in the equilibrium distances between the two Y atoms so that if the original coordinate q be replaced by a new normal coordinate $q^* = q - \delta$, where δ is the amount by which the equilibrium position is shifted, the centrifugal correction terms linear in q^* will disappear, but new terms, always quadratic in q^*, will appear elsewhere. Such terms can readily be dealt with by the perturbation methods. It may readily be verified that the same result could have been achieved by operating on λH_1 with the transformation function $S = (h^2/8\pi^2 I)[J(J+1) - l^2]\,(2q/a)$, I being the moment of inertia, J and l the quantum numbers of total and internal angular momentum, respectively, and a one of the anharmonic constants.

The problem of linear polyatomic molecules has been further dealt with by Sayvetz (1939), particularly with regard to the X_2Y_2-type molecule. In a very recent paper this same type of molecule has been treated quite generally to second order by Wu (1940), who has extended the work of Sayvetz and employed the techniques invented by Adel and Dennison in dealing with the XY_2-type molecule. When the results of this calculation are applied to the data which have accumulated on the acetylene molecule, it is possible for Wu to evaluate many of the constants occurring in his equations for the energies.

The non-linear XY_2-type molecule has also been the subject of a quantum-mechanical treatment similar to that described above. The problem was first discussed by Shaffer and Nielsen (1939), who derived the complete quantum-mechanical Hamiltonian and carried out a perturbation calculation to second order of approximation exactly as indicated in (6*b*) without carrying out the transformation (7):‡ The method consists in calculating the second-order energy matrix after the method of Wang for the asymmetric top. The complete energy matrix is a step matrix with a step for each value of the vibrational quantum number. Each step matrix is really the matrix of a semi-rigid rotator with effective moments of inertia that depend linearly upon the vibration quantum numbers

† This is in reality not true since an accidental degeneracy between ν_1 and $2\nu_2$ produces a first-order perturbation. This phenomenon has been discussed in an earlier paper by Fermi (1931) and discussed more fully again by Dennison (1932).

‡ Very recently Shaffer has re-evaluated the rotation-vibration energies of the non-linear XY_2-type molecule using the transformation THT^{-1}. The result is identical with those described above.

and is diagonal in all quantum numbers save K. The only elements which are non-vanishing are the following: $(K|K)$, $(K|K\pm2)$ and $(K|K\pm4)$, so that the matrix is like the one referred to in § 1 except that a new set of subdiagonals has been introduced by the higher approximation. The $(K|K)$ and $(K|K\pm2)$ elements are essentially the same as the corresponding elements for the rigid top (given in equation (4)) except that they contain additional terms quartic in the rotational quantum numbers arising out of the centrifugal forces. The $(K|K\pm4)$ elements, which are new, originate entirely with the centrifugal forces and are quartic in the quantum numbers J and K. The energies are explicitly available, as in the zeroth-order case, by setting the secular determinant $|(K|H|K')-E\delta_{KK'}|$ equal to zero and solving it. The work herein referred to verifies the work of Bonner (1934) on the vibration energy of such molecules and yields explicit relations for the effective moments of inertia and the coefficients of the centrifugal stretching terms. By expansion of the above secular determinant it is possible to obtain a set of relations for the energies of a vibration-rotation state. When these have been determined experimentally they may be identified with the foregoing relations and the constants occurring in them determined empirically.

Randall, Dennison, Ginsberg and Weber have pointed out that the centrifugal stretching is not adequately compensated for by the addition of terms quartic in J and K, since such terms correspond in reality only to the first-order expansion of the contributions to the energy arising out of the centrifugal stretching as computed by them on the basis of their simple model. It is therefore their contention that expressions for the energy computed in the above manner may be considered valid only for fairly small values of J. The author has nevertheless used these relations in an analysis of a new set of completely resolved data on the water-vapour fundamental ν_2 (Nielsen, 1940). The coefficients of the centrifugal stretching terms were evaluated by fitting the theoretical equations for the energies to the empirically determined values in the normal state. To this approximation these coefficients are independent of the vibrational quantum numbers and therefore have the same values in the state ν_2. From the empirically determined levels corresponding to $J=0, J=1$ and $J=2$, the remaining parameters for the state ν_2 were evaluated. With these values inserted in the theoretical equations it was found possible to predict the positions of other vibration-rotation states to high fidelity as far, at least, as $J=6$.

In later work by Darling and Dennison (1940) the calculations on the effective moments of inertia have been repeated and found substantially to verify the above results. These authors have pointed to an interesting relation existing between the effective moments of inertia, namely, that $C_e-A_e-B_e=\Delta$ is independent of the anharmonic constants in the potential energy and may therefore be calculated from the normal frequencies and the shape of the molecule. It is the belief of these authors that this constitutes an important method of verifying that the best values of the moments of inertia have been attained.

Another polyatomic model which has been considered in the above manner,

at least for a large number of vibration states, is the tetrahedrally symmetric molecule XY_4. This is contained in two papers by Shaffer, Nielsen and Thomas, the first dealing with the quantum-mechanical equation and the eigenvalues and functions for the states $V_1\nu_1$, $\nu_2, V_1\nu_1+\nu_{3,4}$ and $\nu_2+\nu_{3,4}$, the second dealing with the application of the methods of vector addition in quantum mechanics to the states $V_3\nu_3$, $V_4\nu_4$ and $V_3\nu_3+V_4\nu_4$. As in the work dealing with the XY_3-type molecule, explicit energy relations may be arrived at by solution of the secular determinant $|(K|H|K')-E\delta_{KK'}|=0$, where the elements are diagonal in all quantum numbers, save K, the non-vanishing ones, as before, being the $(K|K)$, $(K|K\pm 2)$, $(K|K\pm 4)$ elements. These results are tantamount to saying that, in general, the XY_4-type molecule is not a spherical top, but in reality an asymmetric top where the effective moments of inertia A_e, B_e and C_e are all unlike. To what extent this asymmetry becomes important depends upon the nature of the interactions between rotation and vibration. As Jahn (1938b) has indicated, there exists a Coriolis resonance interaction between the normal frequencies ν_2 and ν_4 large enough in the molecules CH_4, SiH_4 and GeH_4 to induce an asymmetry of the molecules in these states so great that the rotation lines are observed to possess a multiplet structure. A similar effect is predicted by the theory for the frequency ν_3, but to a much less extent. The effect there is so small that it cannot be said to have been observed with certainty. Only in the normal state and in such states as V_1, ν_1 (and in ν_2 when it is far removed from ν_4) etc., where the spherical symmetry is preserved (in which cases the $(K|K\pm 2)$ elements vanish) is the molecule a spherical top, and in these states only where $J=0$ and $J=1$. In states of higher J value, a splitting of the levels due to the $(K|K\pm 4)$ elements, originating with the centrifugal distortion, sets in. Very recent measurements of Dooley (1940) on the frequency ν_4 in methane and the new measurements now being carried out on the spectra of SiH_4 and GeH_4 in this laboratory should present a real opportunity for the application of the theory.

Another model to which similar attention has been given is the ZXY_3 type, for which Dennison and Slawsky (1939) have computed the centrifugal distortion in the normal state. In this state the interactions are at a minimum, particularly the Coriolis interactions which reduce to zero, so that the calculations are here relatively simple. Dennison and Slawsky have computed only the diagonal elements of the energy matrix and have considered as negligible the elements $(K|K\pm 4)$, which also must arise from the centrifugal distortion. As in the problem just discussed, such elements will cause a splitting in the rotational energy levels for values of J greater than $J=1$. A re-measurement of the bands in the spectrum of methyl iodide is being made by Lagemann (1940) in this laboratory. These measurements reveal that the Q-branches in the perpendicular bands are definitely complex in structure. The appearance of these is essentially that of a complex doublet structure such as might be expected from the presence of $(K|K\pm 4)$ elements in the energy matrix. The complete problem of the interaction between rotation and oscillation for any vibration state has been studied

by Shaffer,* and will be published soon, in its application to the spectra of such molecules.

It may perhaps be regarded as an ambitious aim, but to the author it appears that, with the exception of certain molecular models of anomalous behaviour, it might be possible to treat the problem of the interactions between rotation and vibration to the above order of approximation for a perfectly general molecule of N atoms. Inspection of some of the particular cases referred to in the above works will show that the energy relations computed to second order of approximation contain certain constant coefficients and parameters which depend upon the normal frequencies, the size and shape of the molecule and the anharmonic constants, all of which are related in a perfectly definite manner to the coefficients of the various terms which occur in the Hamiltonian operator. If now a convenient notation is adopted there is no more fundamental difficulty encountered in setting up this operator for a molecule consisting of N atoms than for a molecule of five atoms. Moreover, the transformation THT^{-1} can be carried out as readily for a general case as for a particular one. The solution of the zeroth-order problem will be entirely analogous to the solution for a particular case, and the perturbation calculations would probably be only a little more tedious than for such a molecule as the XY_4 model. The final energy relations would again contain parameters and constants all of which would be functions of the normal frequencies, the size and shape of the molecule, etc., and which would bear perfectly definite relationships to the constants occurring in the Hamiltonian operator. For a particular case it would then only be required to determine the normal frequencies, the size and shape of the molecule and the non-vanishing anharmonic coefficients. Once these have been arrived at, the particular form for the energies is determined.

The problem of free rotation in molecules is always a fascinating one. The only case in which actual direct experimental observations have been made on this phenomenon is in the case of the spectrum of methyl alcohol. In a very recent paper by Koehler and Dennison (1940) this problem has been treated in considerable detail, and the results are found to be in agreement qualitatively with the measurements on the spectrum reported by Borden and Barker (1939) in the near infra-red and with measurements now being made in the Michigan laboratory at longer wave-lengths by Lawson and Randall (1930).

The preceding pages have been devoted principally to a discussion of the problem of interactions in polyatomic molecules. In conclusion it is well to underline again what has previously been emphasized, namely, that the background for these theoretical researches is the actual experimental work which has gone before. For example, it was the measurements already made by Cooley (1935) on the spectrum of methane and the work by Bennett and Meyer on the methyl halide spectra which gave stimulus to the theory of Teller, subsequently developed by Dennison and Johnston and others, on the first-order Coriolis effects. These

* Private communication.

effects were first-order effects in the spectrum which could hardly escape the attention of the observer of that decade, working with the relatively small resolving power available to him. The accidental discovery of the complexity of the rotational pattern of the fundamental band ν_4 in the spectrum of SiH_4 stimulated a series of experiments in this laboratory which gave impetus to investigations into the Coriolis perturbation of one vibration frequency upon another and led to the theory of Childs and Jahn for tetrahedrally symmetric molecules and the explanation by Nielsen of the anomalous spacing of the Q-branches in certain formaldehyde bands. The measurements by Randall and his co-workers in the far infra-red on the water-vapour molecules were instrumental in bringing about a study of the centrifugal-distortion interactions in such molecules. These supplementary theories have in turn aroused the interest and curiosity of molecular spectroscopists to the extent that complete and exact theories for at least a limited number of molecular models have been developed, no doubt soon to be followed by others. These developments serve two purposes: first, they are of great aid in the classification of the data already available for such molecules, and secondly, they will frequently point to other and new details which have escaped detection, but which deserve to be hunted for experimentally. In many cases they will even stimulate the experimenter to improve his technique in order to find details which previously had escaped him. For example, the work referred to in the above, all indicates that the centrifugal distortion of the molecules is such as to cause a splitting of many of the rotational energy levels, so that many of the rotational lines should be observed as doublets or probably as multiplets instead of as single lines. This has prompted the reinvestigation in this laboratory of the spectrum of methyl iodide. The result, as we have already point out, is that the Q-branches are in many instances found to be complex or at least double in structure.

The data available adequately to test the interaction theory is, in spite of what has already been done, inadequate and meagre. Really to test these theories not only must more data be gathered, but much better data as well. The achievements of Randall in the far infra-red must be regarded as most outstanding, and they present a definite challenge to experimenters working in the near infra-red. As we have already pointed out, the vacuum spectrograph of Randall has successfully separated lines only 0·25 cm^{-1} apart in the far infra-red. Much of his success is, no doubt, due to the large gratings he has available, the large apertures of his mirrors and in no small way to the optical perfection which he has been able to realize in his instrument.

No revolutionary improvements have been accomplished in the techniques of the near infra-red. In a series of experiments which may be regarded as preliminary to the construction of a vacuum self-recording near-infra-red spectrometer of high resolving power, the writer has attempted, simply by very careful adjustments of his present instrument, to approximate to the practical resolving power realized consistently by Randall at longer wave-lengths. The degree

of success attained is not discouraging and is, perhaps, best visualized by reference to certain unpublished works from this laboratory. The fundamental ν_4 in the methane spectrum has been re-measured again by Dooley, who has consistently separated lines at 7·7 μ which were as little as 0·3 cm^{-1} apart. In a re-measurement of the band spectra of C_2H_2, C_2D_2 and C_2HD, A. H. Nielsen encountered separations as small as from 0·25 cm^{-1} cm. to 0·3 cm^{-1}; and in resolving the bands in the water-vapour spectrum between 2·0 μ and 7·0 μ the writer has recorded many intervals of the same order of magnitude. This type of work demands, of course, very narrow slit-widths, but in addition it is requisite that the data be recorded at very small intervals on the circle. In the investigations just referred to observations were taken at five-second intervals on the spectrometer circle. This makes the measurement of a band a very tedious experience.

At the present time that which actually limits the practical resolving power of infra-red spectrometers for the near infra-red is the size of the galvanometer deflections. To give reproducible details in the spectrum it is, of course, necessary that the actual galvanometer deflections be large compared to the errors introduced by unsteadiness of the recording system. On the other hand, to increase the practical resolving power requires the narrowing down of the spectrometer slits and, as is well known, reducing the size of the slit by a given factor reduces at the same time the amount of radiation falling on the thermocouple by the square of that factor. To be able, therefore, to reduce the slit widths further with advantage will require an increase of the radiation density at the thermocouple. This may be accomplished in two ways: by increasing the sizes of the mirrors and the grating of the spectrometer and by increasing the intensity of the light from the light source.

The size of the grating is limited by the number of lines which it is practical to rule on a surface. The gratings which are used for infra-red spectroscopy are echelette gratings of which the efficiency is determined by the shape of the groove cut by the ruling tool. The number of lines which, in the past, it has been feasible to rule on an echelette grating has been fixed to a large extent by the lifetime of the diamond point, i.e., the number of lines which can be ruled before the shape of the diamond cutting tool is materially altered. In the Michigan laboratory the lifetime of the ruling point has been enhanced by cutting the gratings on solder surfaces and, for the far infra-red, gratings as large as 10×20 inches have been ruled with considerable success. For the near infra-red, where many more lines per inch are required, the size has heretofore been much smaller. It does not appear too visionary, however, to predict that gratings with at least twice, and perhaps four times, the present area will be ruled in the future.

The new technique developed by R. W. Wood at Johns Hopkins University for making replica gratings sounds a new and hopeful note. These replicas are made from a master grating on films which are then cemented, face up, on optical flats. The replica is then covered by evaporation with a coating of aluminium metal. Their optical quality and the reflecting power are very high and, according to the testimony of many experimenters, their resolving power appears to be

almost, or perhaps entirely, as good as that of original rulings now in use. In figure 5 are shown two curves on a group of water-vapour absorption lines (Nos. 84–86 in the Plyler and Sleator (1932) curve for the 3·2 μ region), the first (A) measured with an original grating, the second (B) measured with a replica loaned to the writer by Professor Wood. Both have the same number of lines per inch. While the above curves do not constitute an absolutely acid test for the replica gratings, since the lines shown do not represent the ultimate in resolution obtained in this laboratory with an original ruling, they do attest to their high quality. Professor Wood has even been able to piece together several replicas on one flat, thereby achieving the effect of one large grating. These replicas have the added advantage that they can be produced much less expensively than the original ruling.

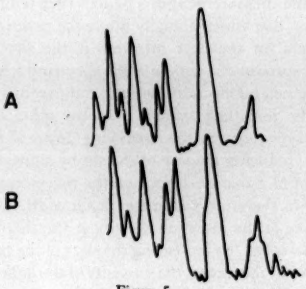

Figure 5.

Until the present time very little improvement has been made in the matter of sources of radiation. The writer is of the belief that at the present time the Nernst glower represents the most satisfactory source of radiation for the near infra-red. The limitation of these is that they "burn out" at the ends where the platinum terminal wires are fastened on. Dr. Earle Ebers, working in this laboratory, found that this difficulty could be circumvented by putting on new terminals in roughly the following manner. The old terminals are completely cleaned off or even broken off. A sufficient amount of platinum wire or foil is then wound around the end to form a small ball. This end of the glower is then held in the flame of an oxygen-gas jet until the platinum just melts and forms a solid ball. It is important that the blast of oxygen should not be too intense since otherwise it may blow the molten platinum off the end of the glower. Into these platinum balls flexible platinum terminal wires can then be welded. Through a glower otherwise rated at 1·0 a., but with the new terminals attached, it has been possible for the writer to pass a current of 2·5 a. for a period of about 30 minutes. When the glower does disintegrate it is because of a boiling off of the glower material itself and not because of a breakdown at the terminals. This transfers the limitation of the source from the terminals to the filament itself.

It appears, therefore, to the writer that it is not beyond the reach of the infra-red spectroscopist to build instruments capable of resolving lines separated by intervals as small as 0·1 cm^{-1}. The necessity of then again reducing the intervals on the spectrometer circle at which data must be recorded virtually demands the introduction of automatic recording.

REFERENCES

ADEL, A. and DENNISON, D. M., 1933. *Phys. Rev.* **43**, 716; **44**, 99.
BARKER, E. F., 1939. *Phys. Rev.* **55**, 658–62.
BENNETT, W. H. and MEYER, C. F., 1928. *Phys. Rev.* **32**, 888–905.
BONNER, L. G., 1934. *Phys. Rev.* **46**, 458.
BORDEN, A. and BARKER, E. F., 1939. *J. Chem. Phys.* **6**, 553.
CHILDS, W. H. J. and JAHN, H. A., 1939. *Proc. Roy. Soc.* A, **169**, 451.
COOLEY, J., 1925. *Astrophys. J.* **62**, 73.
DARLING, B. T. and DENNISON, D. M., 1940. *Phys. Rev.* **57**, 128–39.
DENNISON, D. M., 1931. *Rev. Mod. Phys.* **3**, 280–345.
DENNISON, D. M., 1932. *Phys. Rev.* **41**, 304–13.
DENNISON, D. M. and JOHNSTON, M., 1935. *Phys. Rev.* **47**, 93.
DENNISON, D. M. and SLAWSKY, Z., 1939. *J. Chem. Phys.* **7**, 509.
DOOLEY, D., 1940. *Diss. Ohio State University.*
EBERS, E. and NIELSEN, H. H., 1937. *J. Chem. Phys.* **5**, 822–7.
EBERS, E. and NIELSEN, H. H., 1938. *J. Chem. Phys.* **6**, 311–15.
FERMI, E., 1931. *Z. Phys.* **71**, 250.
FUES, E., 1927. *Z. Phys.* **44**, 91–110.
FUSON, N., RANDALL, H. M. and DENNISON, D. M., 1939. *Phys. Rev.* **56**, 983–1000.
JAHN, H. A., 1938 a. *Proc. Roy. Soc.* A, **168**, 469.
JAHN, H. A., 1938 b. *Proc. Roy. Soc.* A, **168**, 495.
JAHN, H. A., 1939. *Proc. Roy. Soc.* A, **169**, 451.
JORDAHL, O. M., 1934. *Phys. Rev.* **45**, 87.
KOEHLER, J. S. and DENNISON, D. M., 1940. *Phys. Rev.* **57**, 1006.
LAGEMANN, R., 1940. *Diss. Ohio State University.*
LAWSON, J. R. and RANDALL, H. M. (To appear shortly.)
MURPHY, G. M., 1940. *J. Chem. Phys.* **8**, 71–7.
NIELSEN, A. H. and NIELSEN, H. H., 1935. *Phys. Rev.* **48**, 864.
NIELSEN, H. H., 1939. *Phys. Rev.* **55**, 290–3.
NIELSEN, H. H., 1940. *Phys. Rev.* **57**, 346.
PLYLER, E. K. and SLEATOR, W. W., 1931. *Phys. Rev.* **37**, 1493.
RANDALL, H. M., DENNISON, D. M., GINSBURG, N. and WEBER, L., 1937. *Phys. Rev.* **52**, 160.
SAYVETZ, A., 1939. *J. Chem. Phys.* **7**, 383–9.
SHAFFER, W. H. and NIELSEN, H. H., 1939. *Phys. Rev.* **56**, 188–202.
SHAFFER, W. H., NIELSEN, H. H. and THOMAS, L. H., 1939 a. *Phys. Rev.* **56**, 1052–9.
SHAFFER, W. H., NIELSEN, H. H. and THOMAS, L. H., 1939 b. *Phys. Rev.* **56**, 895–907.
STEWARD, W. B. and NIELSEN, H. H., 1935 a. *Phys. Rev.* **47**, 878.
STEWARD, W. B. and NIELSEN, H. H., 1935 b. *Phys. Rev.* **48**, 862.
TELLER, E., 1934. *Hand- und Jahrb. d. Chem. Phys.* **9**.
TELLER, E. and TISZA, L., 1932. *Z. Phys.* **73**, 791.
VAN VLECK, J. H., 1929. *Phys. Rev.* **33**, 467–506.
WEINBERG, A. and ECKART, C. *J. Chem. Phys.* **5**, 517.
WILSON, E. BRIGHT, Jr., 1936. *J. Chem. Phys.* **4**, 313.
WILSON, E. BRIGHT, Jr. and HOWARD, J. B., 1936. *J. Chem. Phys.* **4**, 262.
WU TA-YOU, 1940. *J. Chem. Phys.* **8**, 489.

APPENDIX

As an example of the usefulness of the contact transformation THT^{-1} described in §2 we shall, in this Appendix, compute by means of it the rotation-vibration energy to second order of approximation for a diatomic molecular model which may then be regarded as a prototype of other more complicated models.

The atomic masses will be designated by m_1 and m_2, respectively, and the internuclear axis is taken to coincide with the z axis of a movable system of coordinates x, y and z. In this system of coordinates, the origin of which is made to coincide with the centre of mass of the molecule, the nuclei are free to oscillate with respect to each other along the axis z, their instantaneous coordinates being z_1 and z_2. It is readily shown that the total classical energy of the molecule will be

$$T+U=\tfrac{1}{2}\mu\dot{\eta}^2+2\pi^2\nu_0^2\mu\left[\eta^2-\left(\frac{c}{z_0}\right)\eta^3-\left(\frac{d}{z_0^2}\right)\eta^4+\ldots\right]+\tfrac{1}{2}I_x(\omega_x^2+\omega_y^2), \qquad \ldots\ldots(A\,1)$$

where μ is the reduced mass (m_1m_2/m_1+m_2), $\eta=z_1-z_2$, z_0 is the equilibrium value of η, c and d are constants, ν_0 is the classical frequency of oscillation of the molecule and $I_x=I_y$ is the moment of inertia of the molecule, $m_1z_1^2+m_2z_2^2$.

The Hamiltonian operator H which occurs in the Schrödinger equation $(H-E)\Psi=0$ has been obtained after the method of Wilson and Howard, and when expressed in orders of magnitude we have to second approximation:

$$\left.\begin{aligned} H_0&=\pi\nu_0(p^2+q^2)+(P_x^2+P_y^2)/2I_x^0,\\ H_1&=-(P_x^2+P_y^2)q/(2\pi\nu_0I_x^{03})^{\frac{1}{2}}-(\pi\nu_0/2I_x^0)^{\frac{1}{2}}cq,\\ H_2&=\tfrac{3}{2}(P_x^2+P_y^2)q^2/(2\pi\nu_0I_x^{02})-(d/2I_x^0)q^4. \end{aligned}\right\}\ldots\ldots(A\,2)$$

In the above p is the momentum conjugate to the coordinate q, which has been so chosen that $q=2\pi\nu_0\mu^{\frac{1}{2}}\eta$ and the operator p^2+q^2 has the eigenvalues $(2V+1)\hbar$. P_x^2 and P_y^2 are the x and y components of the operator p^2 which is the square of the total angular momentum. This operator is known to have the eigenvalues $J(J+1)\hbar^2$, and since in the above case the z component of P^2, P_z^2, is identically zero, we may everywhere replace $P_x^2+P_y^2$ by its equivalent P^2, which may be set equal to $J(J+1)\hbar^2$. The energy in zeroth approximation becomes, therefore,

$$E_0=(V+\tfrac{1}{2})\hbar\nu_0+J(J+1)\hbar^2/8\pi^2I_x^0. \qquad \ldots\ldots(A\,3)$$

The Hamiltonian must now be transformed into H' so that $H_0'=H_0$, $H_1'=H_1-i(H_0S-SH_0)$ (which for the diatomic molecule is identically zero) and $H_2'=H_2+(i/2)[SH_1-H_1S]$. With the aid of table 1 in §2, a study of the equations (A 2) reveals that the second requirement above may be accomplished by choosing S, so that

$$S=J(J+1)\hbar p/(8\pi^2\nu_0I_x^{02})^{\frac{1}{2}}+(1/2\pi\nu_0I_x^0)^{\frac{1}{2}}(p^3/3+qpq/2)\frac{c}{\hbar}. \qquad \ldots\ldots(A\,4)$$

After the transformation has been carried out we obtain for H_2'

$$\begin{aligned} H_2'=&-J^2(J+1)2\hbar^4/(8\pi^2\nu_0^2I_x^{03})-[3J(J+1)\hbar^2/4\pi\nu_0I_x^{02}][c(p^2+q^2)+q^2]\\ &-(2I_x^0)^{-1}[c^2(p^2q^2+q^2p^2+qp^2q)+(3c^2/2+d)q^4]. \qquad \ldots\ldots(A\,5) \end{aligned}$$

The contributions to the energy E_2 arising from H_2' will be only its diagonal elements $(V, J\,|H_2'|\,V, J)$. With the aid of the quantum conditions $(qp-pq)=i\hbar\delta$ these may readily be obtained. When E_2 is added to E_0 we obtain the following relation for the energy:

$$E(V,J)=hc\{G(V)+F(J)\}, \qquad \ldots\ldots(A\,6)$$

where

$$G_0(V)=G_0+G_1(v+\tfrac{1}{2})+G_{11}(v+\tfrac{1}{2})^2,$$

in which

$$G_0=(-h/16\pi^2I_x{}^0c)(7c^2/8+3d/4),$$
$$G_1=\nu_0/c,$$
$$G_{11}=(-h/8\pi^2I_x{}^0c)(15c^2/4+3d/2),$$

and where

$$F(J)=B(V)J(J+1)+DJ^2(J+1)^2,$$

in which

$$B(V)=B_0-a(v+\tfrac{1}{2}),$$

B_0 being $(h/8\pi^2I_x{}^0c)$ and $a=+\frac{3}{2}(1+2c)(h^2/16\pi^4I_x{}^{02}\nu^0c)$. The quantity $D=-(4B_e^3/G_1^2)$.

The relations (A 6) for the rotation-vibration energies of a diatomic molecule will be seen to be equivalent to those obtained by Fues in 1926.

An extension of this method is now being carried out by L. H. Thomas of this University and will be published in the near future.

NOTES ADDED IN PROOF

1. Professor Dennison has kindly informed the writer that the measurements of Professor Barker may be so interpreted as to give satisfactory agreement between theory and experiment.

2. Measurements made in this laboratory by Straley and Tindal and Nielsen (1940) establish quite definitely that the less intense maximum occurring in the spectrum of SiH_4 near $10\,\mu$ is due to ν_3. Moreover, in the spectrum of GeH_4, the band reported by Steward and Nielsen is found to be ν_3 in reality and a much more intense band, identified as ν_4, is found near $12{\cdot}0\,\mu$. For SiH_4 it is found that $\nu_3=975$ cm.$^{-1}$ and $\nu_4=910$ cm.$^{-1}$; for GeH_4, $\nu_3=932$ cm.$^{-1}$ and $\nu_4=818$ cm.$^{-1}$

THE GAMMA RADIATIONS EMITTED IN NUCLEAR PROCESSES

By N. FEATHER, Ph.D.

§1. INTRODUCTION

Various aspects of the broad topic of the quantum radiations of atomic nuclei have been examined in previous volumes of these *Reports*—either in the annual surveys covering Atomic Physics generally (Andrade, 1934; Feather, 1935 and 1936; Collie, 1937), or, more recently, in the articles on particular departments of nuclear research (May, 1936; Moon, 1937; Walke, 1939). In this volume, for the first time, a Report is to be devoted specifically to nuclear γ-radiation; it is appropriate, therefore, to preface it by a short summary of the important results bearing on this topic which were obtained in the early days of the science of radioactivity.

It is not, perhaps, generally recognized that it was not until 1922 that the first suggestion of the emission of electromagnetic radiation from atomic nuclei was based upon the results of experiment. The nucleus of the atom had been universally accepted as a physical entity almost as soon as Rutherford explained the phenomenon of large-angle scattering of α particles on this basis in 1911, but it was not until more than ten years later, when Ellis (1922) interpreted the regularities in the natural β-ray spectra of the heavy radioelements with the aid of the idea of quantum radiations describable in terms of schemes of energy levels quite distinct from those appropriate to the characteristic X radiations of the elements, that the most penetrating of the radioactive radiations, first discovered by Villard in 1900, were proved to originate in specifically nuclear processes. Soon a great deal of information concerning the quantum energies of these radiations was collected (Rutherford, Chadwick and Ellis, 1930), and it was shown that they were emitted, presumably in processes of internal reorganization, always as a result of the previous ejection of a disintegration particle (α or β particle) from the nucleus (Meitner, 1925; Ellis and Wooster, 1925). But it was not until 1932 that the first clear evidence of the interrelation of the two processes—the primary process of disintegration and the subsequent act of emission of radiation—was obtained, and the whole phenomenon was exhibited, for α-active substances at least, as exemplifying the principle of energy-conservation in the nuclear domain. In 1929 Rosenblum had discovered the "fine-structure" components of the α radiation of thorium C, and three years later Ellis (1932) definitely established the correctness of a suggestion, due first of all to Gamow (1930), that the emission of a particle with less than the normal energy, in this

case, results always in a product nucleus having excess energy of excitation, which may afterwards be emitted as a quantum, or as successive quanta, of γ radiation. Other examples of the same sort were quickly recognized, and now all α-ray bodies which also emit γ radiation are believed to belong to this class. For nearly all of them the energy spectrum of the α radiation has been investigated by the magnetic focusing method, and the expected fine-structure components have been discovered (Lewis and Bowden, 1934; Rosenblum, Guillot and Perey, 1936). Although, in matters of detail, there is still a great deal for theory to explain, the essential features of the phenomenon are obviously understood (Gamow, 1937); so far nothing in the results of α- and γ-ray measurements runs contrary to the basis of interpretation which was originally proposed.

By comparison, the results of experiments on β-active substances which are also sources of γ radiation are very much less complete. Here, in the simplest case, when no γ radiation is emitted, it is well known that the disintegration particles have all energies from zero up to a maximum characteristic of the body concerned—so that energy conservation appears to fail from the start—but there is growing evidence to show that this maximum energy is the true disintegration energy, and that the failure of the conservation law is more apparent than real (see May, 1936). This was the basis of the bold suggestion concerning the interrelation of β- and γ-ray processes put forward by Ellis and Mott in 1933. From their point of view, the spectrum of the upper energy limits of "partial" continuous spectra, in the β-particle case, corresponds to the discrete fine-structure spectrum of the α particles, already discussed, and the relative numbers of particles belonging to different partial continuous spectra to the intensities of the α-ray "lines". But, clearly, it is no easy task to analyse experimentally a complex continuous spectrum of particles into the partial spectra of which it is believed to be composed. Such an investigation calls for the application of methods of coincidence counting: in respect of the heavy radioelements the initial experiments of this type were made by v. Baeyer (1935) and Bothe and Maier-Leibnitz (1937). In broad outline, these experiments confirmed the suggestions of Ellis and Mott: we shall return to this general method of experiment later in this report.

At a relatively early stage a considerable amount of information of great interest for nuclear theory was provided by a particle-process, connected with the emission of γ radiation, which was observed with the heavy radioelements. This process also has a frequent counterpart, as we shall see, in certain transformations in which light nuclei are involved. We have already mentioned the natural β-ray spectra of the heavy elements, and implied that these are to be interpreted as internal-conversion (photoelectron) spectra. (In some cases, when a quantum of γ radiation might have been emitted, an electron of the extranuclear system escapes from the atom instead, taking up the excess energy provided by the nucleus.) The process with which we are now concerned is one in which a particle escapes from the nucleus itself, as an alternative to the emission of a quantum of radiation. The nuclear species is thereby changed, of course, but

there is accomplished in a single act what would otherwise involve two successive transformations—the emission of a quantum and then a particle (of lower energy than in the one-stage process). The long-range α particles from radium C′ and thorium C′ were first recognized as being emitted by nuclei still in states of excitation following the previous β disintegrations (RaC.C′ and ThC.C′) (Feather, 1929; Fowler, 1930); subsequent investigations have entirely confirmed this general interpretation (Rutherford, Lewis and Bowden, 1933; Ellis, 1934). Since the relation between disintegration constant and energy is reasonably well known in the case of α emission, measurements of relative intensities, in competitive α–γ processes such as these, constitute our most direct means of information regarding the transition probabilities (or lifetimes) which are to be assigned to the radiative processes of nuclei. From considerations of this type it would appear that half-value periods of the order of 10^{-12} sec. are normal for γ rays of quantum energy about 10^6 ev. emitted by heavy nuclei.

Having thus summarized the main items of our earlier knowledge of the quantum radiations of the natural radioelements, for the purposes of the present introduction it remains only to enumerate the various other aspects of more recent research which have to do with nuclear γ-radiation. These we shall have to discuss more fully in the body of this report. In the first place, though it was not the earliest in chronological sequence, we should mention the large amount of work which has been done on the γ radiations of "artificially" produced radioelements. Very soon after the original discovery of induced radioactivity by Curie and Joliot in 1934, Fleischmann (1934) found evidence for the emission of γ rays in certain cases. Nowadays a search for these radiations is part of the normal procedure of investigation. But for the fact that the intensity of the radioactive source is generally much smaller than is the case with the naturally occurring radioelements, the usual methods of experiment are applicable. So far most of the information refers to negative-electron-active species (with positron-active substances there is always the relatively very intense annihilation radiation of the positrons to be allowed for), but it has recently been shown that certain positron-active substances, also, emit nuclear γ-radiation. It would have been very surprising if this had not been so.

The emission of γ radiation as the direct result of the transformation of a light nucleus in collision with a fast-moving α particle was first established by Bothe and Becker in 1930. From experiments described in the previous year, Rutherford and Chadwick (1929) had concluded that, in the transformation of aluminium by α particles, product nuclei are formed with different amounts of intrinsic energy (or with slightly different masses). The detection by Bothe and Becker of γ radiation emitted in this and in other transformations (it was later shown that the penetrating radiations which they observed were, in some cases, in part of a non-quantum character, viz. neutrons) led to the general interpretation which is now universally accepted. We say that as the result of any transformation process, if sufficient energy is available, product nuclei may be formed initially in

states of excitation, in transitions from which γ radiation may subsequently be emitted. During the last eight years the validity of this statement, in respect of transformations produced by neutrons, protons and deuterons, as well as by α particles, has been abundantly verified. It appears that in many cases, too, the nuclei of the atoms of the substance bombarded with particles are raised to states of higher energy-content by a type of non-capture excitation (Savel, 1935; Schnetzler, 1935; Lea, 1935; Barnes and Aradine, 1939; Powell, May, Chadwick and Pickavance, 1940), and it becomes a matter for experiment to distinguish between the γ radiation emitted in the de-excitation of these nuclei and that emitted as the result of any other processes of transformation produced in the same bombardment. Dee and Gilbert (1936) were the first to show that, with a light nucleus (in their case ^{8}Be), de-excitation of a higher energy state might take place predominantly by particle emission (in their case α-particle emission) rather than by the emission of radiation. It now appears very probable that, with heavier nuclei, when the degree of excitation and the general instability are sufficiently great, the emission of neutrons may also compete with the emission of quanta in certain cases (Booth, Dunning and Slack, 1939).

The discovery that "artificially" radioactive species may be produced by the action of neutrons of "thermal" energies on many substances, and that such active species are almost invariably isotopic with the elements bombarded, led Fermi and his collaborators (1934) to conclude that slow neutrons may enter the nuclei of most elements and merely be captured. Clearly, in such a process, γ radiation must be emitted before the product nucleus reaches the state of lowest energy. This γ radiation has been widely studied in recent years. When capture of the incident neutron results in a stable, rather than a radioactive, nucleus—as it must do with many of the elements of even atomic number—the detection of this "capture radiation" is the most direct evidence that the process has taken place at all (Amaldi *et al.*, 1938). The occurrence of radiative capture of protons by nuclei was first suggested, as the only reasonable interpretation of precise experiments, by Cockcroft, Gilbert and Walton in 1935. "Artificially" radioactive substances were found to be produced, under proton bombardment, which could not have been formed otherwise than by simple capture of protons. Further exploration of this phenomenon has, however, depended chiefly upon investigations of the γ radiation emitted in the process of capture. Since the first experiments of Hafstad and Tuve (1935a and 1935b), a large amount of very detailed information has been obtained in this way. This information we shall have to discuss later in this report.

Having in mind the direct estimate of lifetimes already made, from the point of view of immediate experiment we have so far regarded the emission of γ radiation as effectively instantaneous. There is now a large body of evidence to show that this is not always the case. As early as 1930 the idea of metastable states of nuclei was utilized by Fowler (1930) in an attempt to explain the very

high degree of "internal conversion" of one of the γ rays of radium C, but, for this purpose, a γ-ray lifetime certainly not more than a thousand times longer than the normal sufficed. Now we have direct knowledge of radiative processes (of smaller energy), in lighter nuclei, with half-value periods measured in minutes or hours. Information regarding these processes is summarized in one of last year's *Reports* (Walke, 1939). At a later stage in the present report, the possibility of obtaining evidence of radiative processes of intermediate lifetime (10^{-4} sec. to 1 sec., for example) will be briefly reviewed.

§2. EXPERIMENTAL METHODS

It has already been stated that, in general, present-day experiments on nuclear γ-radiation involve the use of active sources much feebler than those available in the classic experiments with the heavy radioelements. For this reason the most exact method of deriving energy values—the study of the natural β-ray spectrum in a magnetic spectrograph—cannot generally be applied : exposures of several millicurie-hours with a highly concentrated source are required for the success of this method, using photographic registration, whereas, by contrast, information has frequently now to be obtained when no more than one microcurie-hour of radiation from a relatively bulky preparation is at the disposal of the investigator. In certain laboratories it is true that this limitation is rapidly being removed by improvements in the performance of high-power apparatus for nuclear bombardment, and in a few instances very successful applications of the method of the natural β-ray spectrum have already been announced (Kalbfell, 1938; Seaborg and Segrè, 1939; Valley and McCreary, 1939; Helmholtz, 1940; Lawson and Cork, 1940), but a great deal has been,and has still to be, done using weaker sources. Briefly this section will review the methods developed for this purpose. Obviously, they will depend to a considerable extent on the mean quantum energy of the radiations concerned: as the quantum energy is increased, the photoelectric effect, the Compton scattering process and the process of pair production become, in turn, the important phenomena underlying the methods of investigation. Probably the most generally used methods are based on the second of these phenomena.

For rough investigations, the absorption of the Compton electrons produced in a thin layer of material may be examined—this is the method introduced by Becker and Bothe in 1932—the absorption being studied by the insertion of foils between two thin-walled counters employed in a coincidence arrangement. More refined investigations are clearly possible if "semi-circular" magnetic analysis of the Compton electrons is possible. Here, again, the earliest success is to be attributed to Bothe (1936), though in an isolated experiment, and instruments of greater resolving power and wider use have more recently been constructed by the Japanese workers (Kikuchi *et al.*, 1939; Watase and Itoh, 1939) and, in England, by Curran, Dee and Strothers (1940). In respect of this method it should be noted that, though it possesses, even in the best case, considerably smaller resolving power than the method of

the natural β-ray spectrum, its use is not restricted to sources which can be prepared in the highest degree of concentration and it has the advantage over the other method that reliable estimates of relative intensities can generally be made, which is not the case when the complicated process of "internal conversion" is in question. Furthermore, the magnetic analysis of the Compton electrons can be applied, as Curran, Dee and Strothers have successfully demonstrated, when the radiations under investigation are those emitted, during the bombardment, in processes of transformation produced by "artificially" accelerated particles. It would appear that this new development of the recoil-electron spectrograph will soon completely replace the more laborious method of expansion chamber and magnetic field, which has been much used by many workers. In spite of careful attempts to standardize the expansion-chamber method for Compton electrons (Richardson and Kurie, 1936; Richardson, 1938 and 1939; Curtis and Richardson, 1940), its low statistical accuracy remains an unavoidable drawback, and many errors in the interpretation of results obtained by its aid have been made in the past.

For investigations of the pairs of electrons produced, in the process of "materialization", by high energy quanta, use of the expansion chamber still has much in its favour. In this case, when a thin foil is employed as scatterer, and a magnetic field is applied, precise information regarding the energy of the quantum responsible is obtained for each pair of electrons observed in the chamber (Crane, Delsasso, Fowler and Lauritsen, 1935; Gaerttner and Crane, 1937a and 1937b; Delsasso, Fowler and Lauritsen, 1937; Halpern and Crane, 1939; Crane, Halpern and Oleson, 1940; Gaerttner and Pardue, 1940). The statistical feature of the investigation remains, but, for a given wastage of experimental material, the resolving power is greater, and the possibility of calculating relative intensities exists as before. This method of investigation obviously makes much fuller use of the data of experiment than that which applies magnetic analysis merely to the positrons belonging to the pairs produced (Alichanov and Dželepov, 1938; Alichanov and Latyshev, 1938). For very weak sources, however, the method of the magnetic spectrograph, rather surprisingly, recovers the advantage, when pair production is studied. Dželepov (1939) has based an ingenious method of analysis on the following considerations. Suppose that a source of monochromatic γ radiation is placed midway between two thin-walled counters (figure 1) in a magnetic field, then, occasionally, when a positron-electron pair is produced in material surrounding the source, and kinetic energy happens to be nearly equally shared between the components of the pair, and the angle of divergence of these components is small, if the field strength is suitably adjusted, the negative electron of the pair will pass through one counter and the positron through the other, and a coincidence will be recorded. For fields of any other magnitude no significant coincidences will be obtained. The essential features of a spectrograph are thus realized: formally, with a complex radiation of quantum energies $E_1, E_2, \ldots E_r, \ldots$, the spectrum of coincidences will exhibit "lines" at energies

such as $\frac{1}{2}(E_\gamma - 2m_0c^2)$. In the matter of intensity, the method is made practicable just because the mean angle between the components of a pair is in general not large, and, as soon as the effective solid angle for detection of particles by either counter is made large enough to take advantage of this fact, the latitude in energy is also such that, at the centre of a spectrum "line", the coincidence rate is not smaller by a considerable factor than the single counting rate in either counter due to electrons produced by the γ ray in question. A preliminary study of a 3-line γ-ray spectrum has been made by this method, using a total "exposure" of the order of 10 microcurie-hours.

Investigations of γ radiations of quantum energy between, say, 10^4 and very nearly 10^6 ev. are best made, with weak sources, by some application of the ordinary photoelectric effect. In general, for the first part of the energy range, from 10^4 to 10^5 ev., no kind of distinction exists between nuclear and extra-nuclear (characteristic x) radiations, and essentially x-ray methods may be

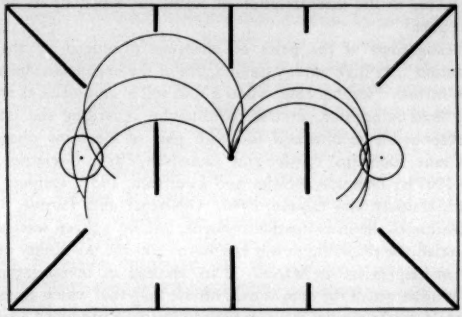

Figure 1.

employed; in the energy range above 10^5 ev. not all of these methods are applicable. Some of the x-ray methods are described in papers by Alvarez (1938) and Abelson (1939), the method of "critical absorption limits" and the "transition-effect" method by the former author (see also Roussinow and Yusephovitch, 1939), and that of the "bent-crystal spectrograph" by the latter. Even in the field of the heavy radioelements, new and precise information has recently been obtained by Ward (1939) by the method of critical absorption. Using a source of ionium of about 2 microcuries effective γ-ray strength, and a high-pressure ionization chamber with a 5 mm. thick aluminium window as detector, Ward clearly demonstrated the emission of a nuclear radiation of 68 ± 1 kev. in about 0·1 per cent of the disintegrations Io. Ra. The method in question depends upon the considerably greater absorption suffered by a given radiation, when its quantum energy lies between the K-absorption energies of two elements of neighbouring

atomic number, in the element of lower atomic number than in that of higher nuclear charge. Thus, when two substances of neighbouring atomic number have been found, such that the element of higher charge is much more transparent, for a given radiation, than the other, it is clear that the quantum energy of the radiation lies between corresponding important X-ray absorption energies of the elements concerned. Obviously this method fails for radiations of quantum energy greater than the K-absorption energy of uranium—$1{\cdot}16 \times 10^5$ ev.; above this energy, for present purposes, its place may be taken by the method of transition effects.

The method of transition effects is concerned with the phenomenon of re-establishment of equilibrium between a primary (γ) radiation and its secon-

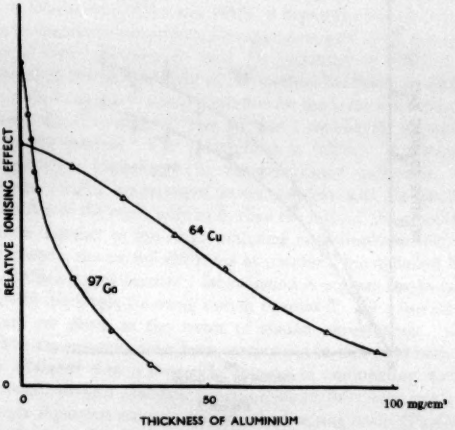

Figure 2.

daries (photoelectrons) when the radiation passes from one absorbing medium to another. Quite clearly, these effects are confined to a thickness of absorber of the same order as the range of the secondary radiations concerned; by adopting a standard procedure, therefore, after suitable calibration, the method may be made to give directly the effective range of the secondary electrons—and in this way the quantum energy of the γ radiation involved. Figure 2 shows the results of Alvarez (1938) relating to the active species ^{97}Ga, for transition from lead to aluminium. From these results, γ radiations of 100 and 250 kev. energy were deduced by a logarithmic treatment of the curve. In figure 3 unpublished results of the writer for the reverse transition process (from aluminium to gold) are shown for the γ radiation of 440 kev. from ^{7}Be. Both in figure 2 and also

in figure 3 calibration curves for pure annihilation radiation (quantum energy 510 kev.), obtained with sources of ^{64}Cu, are given for comparison. Figure 3, in particular, shows the resolving power attainable. In this case ordinates represent the counting rate in a β-particle counter as a percentage of the rate determined with aluminium, only, over the source. It will be remarked that the magnitude of the transition effect, as distinct from the thickness of absorber required for its complete development, is greater, in this range of energies, the smaller the quantum energy of the radiation concerned. Finally, it may be added that each of the curves of figure 3 represents about 40 minutes' counting with a source of approximately 5 microcuries strength.

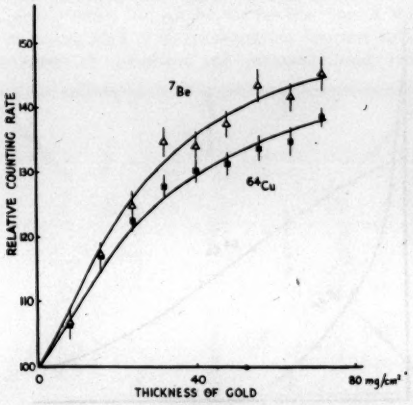

Figure 3.

It has been stated that sources of the order of one microcurie in strength may be successfully investigated by the use of the high-pressure ionization chamber (Ward, 1939). Ward has also developed this method to apply to radiations of intermediate ($0{\cdot}3$ to 3×10^6 ev.) energy, as well as to low-energy radiations (Ward, 1940 b). The method consists essentially in choosing a standard arrangement of source, chamber and absorbers, making use of a large solid angle—and then calibrating this arrangement with monochromatic radiations of various frequencies. An absorption curve is obtained for each radiation (the diminution in ionizing effect in such a case does not follow an exponential law), and afterwards these curves are used for the analysis of the radiations from other substances under investigation. Intensities are deduced by means of a semi-theoretical "ionization function" for the pressure chamber employed, after the validity of this function has been checked using the complex γ rays from a standard source (say Th(B + C)).

Clearly the method cannot be used when the important lines in the γ-ray spectrum are closely spaced—and in any case the absorption cannot be followed through large thicknesses of material (only up to 7 cm. of lead in Ward's arrangement)—but preliminary results have already proved its usefulness in the case of the "artificially" active species ^{7}Be, ^{11}C, ^{13}N, ^{24}Na, ^{27}Mg and ^{56}Mn.

The last general experimental technique to be considered in our discussion of methods is that of coincidence counting. Most obviously, this is a technique which is important for deciding questions of time-correlation for the different radiations characteristic of a given radioelement—and this major use of the technique must eventually be discussed in this report—but, first of all, its employment for the deduction of γ-ray energies and intensities, in favourable cases and with very weak β-active sources, will be referred to. This development is due primarily to Dunworth (1939 a and 1940); it depends upon the use of β- and γ-ray counters in standard coincidence arrangements and upon a knowledge of the efficiencies of these counters, so used, for radiations of different energies. The main problem in the development of the method has been the experimental determination of the {efficiency : energy} relation for the γ-ray counters employed. This having been done, sources of from 10^{-3} to 1 microcurie in strength may be examined by the method. The principle is as follows: if a source emits, possibly amongst other radiations, two time-correlated radiations, A and B, and if two counters, 1 and 2, are arranged (in the simplest case), the one to respond only to radiation A and the other only to B, then the ratio of the coincidence rate for both counters together to the single counting rate simultaneously observed in counter 1 is, clearly, the overall efficiency of counter 2 for radiation B. Similarly, the overall efficiency of counter 1 for radiation A is given by the ratio of the coincidence rate to the single counting rate in counter 2. In a less simple case, mean efficiencies are given as the result of similar experiments. Generally, when coincidence experiments have been carried out in this way, both with two γ-ray counters and also with a β-particle counter in conjunction with a γ-ray counter, and straightforward absorption measurements have been made, in addition, with a single β-particle counter in a standard arrangement (Feather, 1938; Sargent, 1939), it is possible to interpret the crude efficiency values obtained, in terms of a simple energy diagram giving the energies and intensities of the particle and quantum radiations involved. Up to date the method has been standardized by, or successfully applied to, sources of Ra(B + C), Th(B + C), ThC", ^{7}Be, ^{11}C, ^{13}N, ^{24}Na, ^{27}Mg, ^{28}Al, ^{56}Mn, ^{64}Cu, ^{114}In and ^{152}Eu.

The coincidence method has already been, and is being increasingly, employed when the quantum energies of the important γ radiations are sufficiently well known from the start, simply for correlation studies by many workers. Thus, from γ—γ coincidence experiments, the emission of several γ rays in cascade, as a result of the process of neutron capture, has been demonstrated (Aoki, 1937; Fleischmann, 1936 and 1938)—and, in agreement with theoretical predictions, frequently, also, a similar cascade process has been shown to occur with one or

other member of a pair of nuclear isomers (Feather and Dunworth, 1938; Hemmendinger, 1939; Langer, Mitchell and McDaniel, 1939a; Stewart, 1939; Enns, 1939). From $\beta-\gamma$ coincidence experiments, using the β particles issuing from different thicknesses of absorber, energy diagrams have been proposed in respect of the $(\beta+\gamma)$ disintegrations of many artificially produced active species (Norling, 1938 and 1939; Rotblat, 1939; Saha, 1939; Langer, Mitchell and McDaniel, 1939b; Mitchell, Langer and McDaniel, 1940). A more powerful development of this general method of analysis has been described by Feather (1940). In this new arrangement two β-particle counters are employed, on opposite sides of a thin source centrally placed in a magnetic spectrograph, and, with the magnetic field adjusted to bring the electrons of a natural β-ray (photoelectron) line to a focus on one counter, the other counter is made to traverse the complete particle spectrum formed on the other side of the source, so that a "spectrum of coincidences" is obtained. This spectrum clearly has the form of the partial continuous spectrum of those β particles the emission of which is correlated with the emission of the γ ray corresponding to the photoelectron line in question. An estimate of experimental magnitudes indicates that the method may, in favourable cases, be used when the absolute intensity of the photoelectron line is as low as 10^{-2} per disintegration.

So far it has tacitly been assumed that the time-correlated processes studied by the method of concidences occur within so short an interval one of another as to be effectively instantaneous. Actually, a finite time must separate the emission of a particle and a quantum, or of two quanta, from a nucleus, and the (electrical) coincidence method is capable of demonstrating this effect, when the time interval concerned is greater than about 10^{-7} sec. and less than about 1 sec. Dunworth (1939b) and Rotblat (1939) have used this "method of variable resolving time", with two counters, in order to deduce the half-value periods of the short-lived products RaC′ and ThC′ ($\tau = 1{\cdot}45 \times 10^{-4}$ and 2×10^{-7} sec., respectively), and a circuit for use with a single counter—effectively to separate the time-correlated from the random pulses in the counter—has been devised by Dunworth and applied by Ward (1940a) to deduce the periods of ThA, AcA and RaC′ ($\tau = (0{\cdot}158 \pm 0{\cdot}008)$, $(1{\cdot}83 \pm 0{\cdot}04) \times 10^{-3}$ and $(1{\cdot}48 \pm 0{\cdot}06) \times 10^{-4}$ sec., respectively) (see also Roberts, 1940). This method can obviously be applied (though, because of the smaller efficiency of detection, at the expense of longer periods of counting) when γ-ray lifetimes of the same order are in question, and much information is, in future, to be looked for in this direction.

§3. SOME INTERESTING RESULTS

The most interesting features of recent research concerning nuclear γ-radiation are, undoubtedly, on the one hand, the discovery of many instances of radiative processes of long lifetime, and, on the other, the proof that quite frequently a high degree of "internal conversion" occurs, even with elements of relatively low atomic weight. As might be supposed, these two features are intimately related. With heavy radioelements, early direct estimates of γ-ray lifetimes pointed

to values of the order of 10^{-12} sec., and, for all but a few γ rays in respect of which measurement were possible, to quantum yields * greater than 0·9 per nuclear transition. At the same time, a semiquantitative correlation of these facts was also achieved by a successful theory of internal conversion (Hulme, Mott, Oppenheimer and Taylor, 1936). To a first approximation, in this theory, the escape of a quantum and the ejection of an extra-nuclear electron are regarded as competing processes, each contributing its share to the total probability (per unit time) that a given nucleus undergoes de-excitation: for similar processes in lighter nuclei the theory predicts quantum yields even more closely approximating to unity. But all this is based on the explicit assumption that in no case is a change in nuclear angular momentum of more than two units involved in the transition concerned. On the basis of a general result due to Heitler (1936), representation of the emitting nucleus is therefore restricted to the four possibilities, electric and magnetic dipole and quadripole. Naturally, if larger spin changes are contemplated, it becomes necessary to employ multipole fields of higher order in fresh calculations. Now, in the theory of nuclear isomerism put forward by v. Weizsäcker (1936), it is precisely the idea of transitions involving large changes in angular momentum which is developed to account for the facts.† On the basis of this same idea, new calculations concerning internal conversion have been made by Hebb and Uhlenbeck (1938), Dancoff and Morrison (1939), and Nelson and Hebb (1940). In one or two cases it has been possible to subject these new calculations—as well as the revised theory of isomerism itself—to fairly stringent numerical tests. So far the general indications are that the correct explanation of both sets of phenomena has now been achieved.

Broadly speaking there are three measurable quantities by which the theory may be tested: the total transition probability for the metastable state concerned, the quantum yield, and the ratio of the intensities of the K and L conversion electron "lines". Knowledge of these three quantities is equivalent to a knowledge of the absolute probabilities (per unit time) of de-excitation of the nuclear state, by radiation, by K-electron ejection, and by L-electron ejection, respectively. Results obtained with the 6·7-hr. cadmium activity illustrate the points at issue. For this body (which transforms by K-electron capture) Valley and McCreary (1939) have measured the energies and relative intensities of the three β-ray lines, corresponding to conversion in the K, L_I and M levels, of a γ ray of 92·6 kev. energy. Their measurements show that the K and L_I lines in this spectrum are of very nearly equal intensity (in all "normal" cases a K/L_I ratio of about 8:1 is observed). On the basis of the old calculations concerning internal

* In order to avoid the ambiguities attached, by reason of its history, to the use of the term "internal conversion coefficient" in respect of nuclear radiations, the complementary expression "quantum yield" is here borrowed from standard x-ray terminology to express the probability of the escape of a primary quantum from the atom, as a consequence of a nuclear transition involving the release of a given amount of energy.

† In this account some overlapping with §§ 5 and 6 of last year's Report on induced radioactivity (Walke, 1939) is almost unavoidable. The matter common to the two Reports is, however, discussed from different viewpoints in each.

conversion the observed ratio is inexplicable; with the new calculations it can only be explained if a multipole field of high order characterizes the transition between the nuclear states responsible for the radiation in question (Nelson and Hebb, 1940). On the other hand, for such a field to be important, a large difference of spin should also exist between the nuclear states— and the upper state should then be metastable, and the quantum yield should be low. Recent investigations by Alvarez, Helmholtz and Nelson (1940) provide full confirmation of this reasoning. These authors have shown that a half-value period of 40 sec. and a quantum yield of only 2 per cent are to be ascribed to the 92·6 kev. transition. Moreover, they point out that both the K/L_I conversion ratio and the absolute value of the quantum yield require a radiation field of multipole order 4 for their explanation, and that, whilst the theoretical value of the lifetime of the state for a spin change of four units is 30 sec., periods 10^4 times greater or less than this would be expected if the actual value of the spin change were different by one unit in either direction. In this case, therefore, very convincing numerical support is provided for current theoretical ideas.*

By way of a summary, table 1 includes all the information at present available concerning lifetimes and energies of those metastable states of nuclei which transform by the emission of a quantum or the ejection of an extra-nuclear electron, rather than by radioactive (particle) decay. In the present connection, clearly all that can be said, regarding the metastable radioactive species omitted from the table, is that the partial probability of de-excitation of the metastable state is, with them, very much smaller than the reciprocal of the lifetime which is observed. With some nuclei, of course, radioactive decay and de-excitation may both occur, but it appears extremely probable that in general one mode or the other will greatly predominate (cf. $UX_2 - UZ$) (Feather and Bretscher, 1938). In the present state of our knowledge it is not to be expected that all cases investigated should as yet be equally well understood, and here attention may be directed to the results of Valley and McCreary (1939) with ^{80}Br, where, in spite of close study, some obscurity still remains. It appears that two low-energy radiations, each of low quantum yield, are emitted from this body. A highly absorbable but apparently extremely short-lived electron radiation from gadolinium (Amaldi and Rasetti, 1939; Alchazov, Gurevitsch, Kurtchatov and Rukavischnikov, 1939) also seems difficult to explain on current ideas.

It is always interesting to examine any fresh development in the theory of nuclear processes for its relevance to the heavy radioelements, more detailed investigations of the complex radiations from which have frequently proved fruitful of new surprises (Perey, 1939; Perey and Lecoin, 1939). It is not improbable that the developments just discussed have an important bearing in this respect also. The theory of α distintegration, as extended by Gamow, allows certain deductions concerning the spins of excited states to be made from the relative intensities of the lines in the fine-structure spectra of α particles, and the

* Alternative suggestions by Sachs (1940) do not appear to allow of a sufficient range of possibilities to cover all actual cases.

Table 1. (Symbols in italics indicate stable species)

Nuclear species	Half-value period	γ-ray kev. energy	References
$^{49}_{20}$Ca	30 m.	?	Walke, Thompson and Holt, 1940
$^{44}_{21}$Sc	52 h.	250	Walke, 1940
$^{69}_{30}$Zn	13·8 h.	470	Livingood and Seaborg, 1939; Sagane *et al.*, 1939; Kennedy *et al.*, 1939
$^{79/81}_{34}$Se	1 h.	98	Langsdorf and Segrè, 1940
$^{80}_{35}$Br	4·5 h.	48·9 { 37·1 or 25·3 }	Segrè, Halford and Seaborg, 1939; de Vault and Libby, 1939; Siday, 1939; Roussinow and Yusephovitch, 1939; Valley and McCreary, 1939
$^{83}_{36}$*Kr*	1·9 h.	49	Langsdorf and Segrè, 1934
$^{87}_{38}$*Sr*	2·7 h.	370	Du Bridge and Marshall, 1939 and 1940
$^{87}_{39}$Y	14 h.	~500	Du Bridge and Marshall, 1940
$^{89}_{40}$Zr	4·5 m.	?	Du Bridge and Marshall, 1940
$^{99/101}_{43}$Ma	6·6 h.	136, 180	Seaborg and Segrè, 1939
$^{104}_{45}$Rh	4·37 m.	80	Crittenden and Bacher, 1938; Pontecorvo, 1938 a; Crittenden, 1939
$^{107/109}_{47}$*Ag*	40 s.	92·6	Alvarez, Helmholtz and Nelson, 1940
$_{48}$*Cd*	50 m.	?	Dodé and Pontecorvo, 1938
$^{113}_{49}$*In*	1·75 h.	393	Barnes, 1939; Lawson and Cork, 1940.
$^{114}_{49}$In	50 d.	192	Lawson and Cork, 1940
$^{115}_{49}$*In*	4·5 h.	338	Goldhaber, Hill and Szilard, 1939; Barnes and Aradine, 1939; Collins *et al.*, 1939; Pontecorvo and Lazard, 1939; Lawson and Cork, 1940
$^{127}_{52}$Te	90 d.	~125	Seaborg, Livingood and Kennedy, 1940
$^{129}_{52}$Te	32 d.	~100	Seaborg, Livingood and Kennedy, 1940
$^{131}_{52}$Te	1·2 d.	~165	Seaborg, Livingood and Kennedy, 1940
$_{82}$*Pb*	1·6 m.	?	Waldman and Collins, 1940

measured intensities of these " lines " give directly the primary excitations of the states in question. Here the case of ionium, already mentioned in this report, is of particular interest. With a body of such long lifetime, the accurate study of the α-particle spectrum is a matter of great difficulty (Soddy, 1932), but if Ward's estimate (Ward, 1939) of 10^{-3} quantum per disintegration represents even approximately the full degree of excitation of the state of 68 kev. energy, then a very large

difference of spin must separate this state from the ground state of the following nucleus. It will be interesting to see whether, in fact, a nuclear transition of relatively long period is involved. A survey of the whole body of accepted data indicates that similar interest attaches to further studies of the 31·5 kev. state excited in the disintegration Rd Ac . AcX, and possibly of other excited states occurring with bodies of the actinium active-deposit (Kara-Michailova, 1938), as well as of the state of 184 kev. energy in Ra . Rn. Attention has previously been directed to this general question, though with less concern for detail, by Feather and Bretscher (1938) and by Pontecorvo (1938 b and 1939).

The last aspect of research on nuclear γ-radiation for which there is space in this report is that which deals with the processes of capture of protons and neutrons by nuclei. The occurrence of these processes of simple capture was recognized at an early stage in the work on induced radioactivity, and some reference has already been made to the original investigations in earlier reports in this series (Feather, 1935 and 1936; May, 1936; Moon, 1937; Walke, 1939). We shall be concerned here with the more precise studies, recently carried out, of the dependence of capture probability on particle energy (determination of "excitation functions"), and with experiments on the spectral distribution of the capture radiation.

For a general understanding of the process of particle capture the concept of the intermediate nucleus is of prime importance (Bohr, 1936; Breit, 1940). Capture is possible only when the state in which the intermediate nucleus is formed is sufficiently long-lived for the emission of radiation to be an important mode of de-excitation (otherwise expressed, if particle emission from an intermediate state is relatively rare, then the de-excitation probability is effectively that of a radiative process, and the state must obviously be relatively long-lived). Now the principle of indeterminacy requires that a long-lived state shall be one of well-determined energy; thus capture processes must be expected to show all the phenomena of resonance. This is the argument on a general theoretical basis; on the other hand, the resonant nature of capture processes was discovered before the theoretical picture was fully developed (Hafstad and Tuve, 1935 a; Tillman and Moon, 1935; Moon and Tillman, 1936). Recent work, as already implied, has greatly extended our knowledge of these resonance effects, and exhibited much more clearly their great complexity. With protons this work has been done largely by Herb and his collaborators in America (Bernet, Herb and Parkinson, 1938; Plain, Herb, Hudson and Warren, 1940), and in the High Voltage laboratory in Cambridge (Dee, Curran and Petržílka, 1938; Curran, Dee and Petržílka, 1939; Curran and Strothers, 1939). As an instance of the extent of the agreement reached, the following values of proton energies, at resonance, may be given, in respect of the bombardment of fluorine. Values obtained by the American workers are given first, with the slightly less precise Cambridge values in brackets. Over the range of energy studied by both sets of experimenters (0 to 10^6 ev.), the agreement may be said to be complete: 0·334 (0·33), 0·479 (0·47), 0·589 (0·59), 0·660 (0·67), 0·862 (0·87), 0·927 (0·92), 1·26, 1·335, 1·363, 1·67 Mev.

Up to date, resonant features have been found in the energy dependence of the intensity of the γ radiation obtained in the bombardment of the following elements: Li, Be, B, C, F, Na, Mg, Al, P, Cl, and with all these elements, except Cl, these features have been definitely assigned to the isotope or isotopes responsible. For all the elements listed, studies of the γ radiation have been made—at least with the degree of detail which the "secondary-electron coincidence-absorption method" (*v. sup.*) admits. If this method is standardized both in terms of absorption end-point and half- (or quarter-) value thickness, some indication of the spectral constitution of the radiation may be obtained (Rasetti, 1935). Taking the results as a whole, it appears to be the exception, rather than the rule, for the full energy of excitation of the intermediate nucleus (energy of binding of the proton in the final product) to be emitted as a single quantum of radiation. One-quantum de-excitation is apparently predominant for capture by ^{7}Li and ^{13}C, but, with other nuclei, though it also occurs, this process seems to represent a rather rare mode of de-excitation. With ^{26}Mg it is possible that the one-quantum process does not take place to any appreciable extent. Usually, in these capture transformations, the energy available for radiation—as a single quantum, or a number of quanta emitted in succession—is of the order of 10^{7} ev. So far, the results obtained agree completely with expectations based upon known differences in the mass-energies of nuclei, except in the case of fluorine. The resonant transmutation of fluorine by protons has required particularly detailed study before any satisfactory interpretation of results has been achieved, and, even yet, some aspects of this transmutation are not completely understood. When the quantum energy of the radiation was first compared with the energy available in the simple capture process, in this case, it was thought that a clear instance of two-quantum de-excitation was involved: roughly homogeneous radiation of about 6 Mev. was observed, whereas about 13 Mev. of mass energy was available. Dee, Curran and Strothers (1939) were the first to demonstrate that this assumption was untenable; they failed to find the coincidences which would have been expected with two γ-ray counters if the assumption had been correct, and they showed that an increase in the energy of the proton producing the transmutation was without effect on the energy of the quantum emitted. They explained these results—and the then recent discovery of Burcham and Smith (1939) that a group of short-range α particles was emitted, in conditions of resonance similar to those obtaining for γ-ray emission—on the assumption that the resonant process concerned was primarily one of particle (not quantum) release (see also Maclean, Becker, Fowler and Lauritsen, 1939):

$$^{19}_{9}\mathrm{F} + ^{1}_{1}\mathrm{H} \rightarrow ^{16}_{8}\mathrm{O}^{*} + ^{4}_{2}\mathrm{He}.$$

On the basis of this explanation the excited state of ^{16}O formed in the resonance transmutation is one of 6·3 Mev. energy of excitation. Burcham and Devons (1939) confirmed this interpretation by showing that the energy of the short-range α particles (rather than the energy of the quantum) increased by the correct amount when the energy of the protons was raised from one (resonance) value to

another. For the first time, then, clear evidence had been provided of resonant features in a (p, α) disintegration, though the frequent occurrence of such features in (α, p) transmutations had been known for a long time. When this first result had been obtained, Burcham and Devons (1939) were able to show, by further careful investigation, that weak resonant features also characterized the competing (p, α) disintegration with fluorine which leads to the final nucleus, ^{16}O, in the ground state. These less well-marked resonances, however, did not coincide, on the energy scale, with those obtaining in the other process in which γ radiation is emitted, and, in this connection, it should be remembered that the normal transformation $^{19}F(^1H, ^4He)^{16}O$ had been studied for many years without these weak resonances having been detected at all (see, for example, Henderson, Livingston and Lawrence, 1934).

The most recent contribution to the study of the complicated processes of proton-induced transmutation with fluorine is the surprising discovery of Fowler and Lauritsen (1939) that positron-electron pairs are produced with a total energy (? including mass-energy) of $5{\cdot}9 \pm 0{\cdot}5$ Mev., and that the excitation function for this emission also shows resonant features, some resonances being common to this and the γ-ray process, and some peculiar to each. A complete explanation does not at present appear to have been offered, but Oppenheimer and Schwinger (1939) at least conclude that the simple process of internal conversion by pair production is not in question. One suggestion of these authors, that pair-production resonances might coincide with full-range α-particle resonances, seems not improbable on a comparison of the relevant experimental results (Burcham and Devons, 1939; Fowler and Lauritsen, 1939).

With lithium, certain previous discrepant results concerning the γ radiation obtained on proton bombardment have recently been resolved by the discovery by two different sets of experimenters, that, for protons of energy greater than 10^6 ev., non-capture excitation of the nucleus ^7Li (in the state of 440 kev. excess energy) begins to compete seriously with the simple-capture process in which γ rays of 17·5 mev. energy are emitted (Fowler and Lauritsen, 1939; Hudson, Herb and Plain, 1940).

The simple capture of neutrons by nuclei, appreciable with the sources at present available only when the energy of the neutron has been reduced below a few hundred volts, has been described, in its main features, by Moon in vol. 4 of these *Reports*. The resonant nature of the process is well marked, from a qualitative aspect, by the occurrence of the phenomenon of selective absorption with slow neutrons (Tillman and Moon, 1935), but methods have yet to be developed for any very close study of the dependence of capture probability on energy over the important range. At present the method of absorption in boron (it is supposed that the absorption coefficient for slow neutrons in boron varies inversely as the velocity of the neutrons) provides the bulk of our information: the development of more precise methods is essentially concerned with the development of methods of producing, or selecting for use, monokinetic neutrons

in the low-energy range. Various attempts in this direction have already been reported, but the difficulties have been considerable, and progress has in general been slow (Fertel *et al.*, 1940; Baker and Bacher, 1940).

Concerning the spectral distribution of the capture radiation, investigations have been made exclusively by the coincidence-absorption method, the most complete series of observations being due to the Japanese workers (Kikuchi, Husimi and Aoki, 1936). Even more frequently than in the capture of protons, it appears that de-excitation of the initial state of the intermediate nucleus occurs with the successive emission of at least two quanta of radiation. Moreover, in its broad aspects, this feature of the process has been shown to depend on the nuclear charge: if the ratio, f, of the mean quantum energy of the capture radiation to the total energy available in the process, be plotted against the atomic number, Z, a pronounced maximum ($f \sim 0{\cdot}8$) is found for $Z \sim 25$. As already mentioned, with certain elements simultaneous emission of several quanta of capture radiation has been demonstrated by the method of coincidences between γ-ray counters (Fleischmann, 1936 and 1938; Aoki, 1937). When de-excitation takes place by the emission of several quanta in this way, the question arises whether or not the successive transitions are the same, into whichever resonance level the slow neutron is captured. Pontecorvo (1936) has obtained some evidence for differences in the spectral distribution of the capture radiation with gold when the effective capture level is changed by filtration of the incident neutrons. The final result of the process of capture can be studied differentially when active isomers are formed in the process. When this is the case, different sequences of de-excitation must obviously occur, and the ratio of the intensities of production of the isomeric forms can be studied as a function of the energy of the neutrons employed. Preliminary studies of this kind have been made by Fleischmann (1937) and also by Snell (1937) for ^{80}Br, and by Mitchell and Langer (1938) for ^{116}In.

REFERENCES

Abelson, 1939. *Phys. Rev.* **56**, 753.

Alchazov, Gurevitsch, Kurtchatov and Rukavischnikov, 1939. *C.R. Acad. Sci., U.R.S.S.*, **25**, 111.

Alichanov and Dželepov, 1938. *C.R. Acad. Sci., U.R.S.S.*, **20**, 113, 115.

Alichanov and Latyshev, 1938. *C.R. Acad. Sci., U.R.S.S.*, **20**, 429.

Alvarez, 1938. *Phys. Rev.* **54**, 486.

Alvarez, Helmholtz and Nelson, 1940. *Phys. Rev.* **57**, 660.

Amaldi *et al.*, 1935. *Proc. Roy. Soc.* A, **149**, 522.

Amaldi and Rasetti, 1939. *Ricerca Scientifica*, **10**, 115.

Andrade, 1934. *Rep. Progr. Phys.* **1**, 269.

Aoki, 1937. *Proc. Phys.-Math. Soc. Japan*, **19**, 799.

v. Baeyer, 1935. *Z. Phys.* **95**, 417.

Baker and Bacher, 1940. *Phys. Rev.* **57**, 351 A, 1076 A.

Barnes, 1939. *Phys. Rev.* **56**, 414.

Barnes and Aradine, 1939. *Phys. Rev.* **55**, 50.

Becker and Bothe, 1932 a. *Naturwissenschaften*, **20**, 349.

BECKER and BOTHE, 1932 b. *Z. Phys.* **76**, 421.
BERNET, HERB and PARKINSON, 1938. *Phys. Rev.* **54**, 398.
BOHR, 1936. *Nature, Lond.*, **137**, 344.
BOOTH, DUNNING and SLACK, 1939. *Phys. Rev.* **55**, 876.
BOTHE, 1936. *Z. Phys.* **100**, 273.
BOTHE and BECKER, 1930. *Naturwissenschaften*, **18**, 705.
BOTHE and MAIER-LEIBNITZ, 1937. *Z. Phys.* **104**, 604.
BREIT, 1940. *Phys. Rev.* **58**, 506.
DU BRIDGE and MARSHALL, 1939. *Phys. Rev.* **56**, 706.
DU BRIDGE and MARSHALL, 1940. *Phys. Rev.* **58**, 7.
BURCHAM and DEVONS, 1939. *Proc. Roy. Soc.* A, **173**, 555.
BURCHAM and SMITH, 1939. *Nature, Lond.*, **143**, 795.
COCKCROFT, GILBERT and WALTON, 1935. *Proc. Roy. Soc.* A, **148**, 225.
COLLIE, 1937. *Rep. Progr. Phys.* **4**, 173.
COLLINS, WALDMAN, STUBBLEFIELD and GOLDHABER, 1939. *Phys. Rev.* **55**, 507.
CRANE, DELSASSO, FOWLER and LAURITSEN, 1935. *Phys. Rev.* **48**, 100, 102.
CRANE, HALPERN and OLESON, 1940. *Phys. Rev.* **57**, 13.
CRITTENDEN, 1939. *Phys. Rev.* **56**, 709.
CRITTENDEN and BACHER, 1938. *Phys. Rev.* **54**, 862.
CURRAN, DEE and PETRŽÍLKA, 1939. *Proc. Roy. Soc.* A, **169**, 269.
CURRAN, DEE and STROTHERS, 1940. *Proc. Roy. Soc.* A, **174**, 546.
CURRAN and STROTHERS, 1939. *Proc. Roy. Soc.* A, **172**, 72.
CURTIS and RICHARDSON, 1940. *Phys. Rev.* **57**, 1121.
DANCOFF and MORRISON, 1939. *Phys. Rev.* **55**, 122.
DEE and GILBERT, 1936. *Proc. Roy. Soc.* A, **154**, 279.
DEE, CURRAN and PETRŽÍLKA, 1938. *Nature, Lond.*, **141**, 642.
DEE, CURRAN and STROTHERS, 1939. *Nature, Lond.*, **143**, 759.
DELSASSO, FOWLER and LAURITSEN, 1937. *Phys. Rev.* **51**, 391, 527.
DODÉ and PONTECORVO, 1938. *C.R. Acad. Sci., Paris*, **207**, 287.
DUNWORTH, 1939 a. *Nature, Lond.*, **143**, 1065.
DUNWORTH, 1939 b. *Nature, Lond.*, **144**, 152.
DUNWORTH, 1940. *Rev. Sci. Instrum.* **11**, 167.
DŽELEPOV, 1939. *C.R. Acad. Sci., U.R.S.S.*, **23**, 24.
ELLIS, 1922. *Proc. Roy. Soc.* A, **101**, 1.
ELLIS, 1932. *Proc. Roy. Soc.* A, **136**, 396.
ELLIS, 1934. *Proc. Roy. Soc.* A, **143**, 350.
ELLIS and MOTT, 1933. *Proc. Roy. Soc.* A, **141**, 502.
ELLIS and WOOSTER, 1925. *Proc. Camb. Phil. Soc.* **22**, 844.
ENNS, 1939. *Phys. Rev.* **56**, 872.
FEATHER, 1929. *Phys. Rev.* **34**, 1558.
FEATHER, 1935. *Rep. Progr. Phys.* **2**, 74.
FEATHER, 1936. *Rep. Progr. Phys.* **3**, 66.
FEATHER, 1938. *Proc. Camb. Phil. Soc.* **34**, 599.
FEATHER, 1940. *Proc. Camb. Phil. Soc.* **36**, 224.
FEATHER. Unpublished.
FEATHER and BRETSCHER, 1938. *Proc. Roy. Soc.* A, **165**, 530.
FEATHER and DUNWORTH, 1938. *Proc. Roy. Soc.* A, **168**, 566.
FERMI *et al.*, 1934. *Ricerca Scientifica*, **5 (2)**, 282.
FERTEL *et al.*, 1940. *Proc. Roy. Soc.* A, **175**, 316.
FLEISCHMANN, 1934. *Naturwissenschaften*, **22**, 434.
FLEISCHMANN, 1936. *Z. Phys.* **103**, 113.
FLEISCHMANN, 1937. *Z. Phys.* **107**, 205.
FLEISCHMANN, 1938. *Z. Phys.* **111**, 281.

FOWLER, 1930. *Proc. Roy. Soc.* A, **129**, 1.
FOWLER and LAURITSEN, 1939. *Phys. Rev.* **56**, 840.
GAERTTNER and CRANE, 1937 a. *Phys. Rev.* **51**, 49.
GAERTTNER and CRANE, 1937 b. *Phys. Rev.* **52**, 582.
GAERTTNER and PARDUE, 1940. *Phys. Rev.* **57**, 386.
GAMOW, 1930. *Nature, Lond.*, **126**, 397.
GAMOW, 1937. *Structure of atomic nuclei and nuclear transformations*, chap. 6.
GOLDHABER, HILL and SZILARD, 1939. *Phys. Rev.* **55**, 47.
HAFSTAD and TUVE, 1935 a. *Phys. Rev.* **47**, 506.
HAFSTAD and TUVE, 1935 b. *Phys. Rev.* **48**, 306.
HALPERN and CRANE, 1939. *Phys. Rev.* **55**, 260.
HEBB and UHLENBECK, 1938. *Physica*, **5**, 605.
HEITLER, 1936. *Proc. Camb. Phil. Soc.* **32**, 112.
HELMHOLTZ, 1940. *Phys. Rev.* **57**, 248 A.
HEMMENDINGER, 1939. *Phys. Rev.* **55**, 604 A.
HENDERSON, LIVINGSTON and LAWRENCE, 1934. *Phys. Rev.* **46**, 38.
HUDSON, HERB and PLAIN, 1940. *Phys. Rev.* **57**, 587.
HULME, OPPENHEIMER, MOTT and TAYLOR, 1936. *Proc. Roy. Soc.* A, **155**, 315.
KALBFELL, 1938. *Phys. Rev.* **54**, 543.
KARA-MICHAILOVA, 1938. *Proc. Camb. Phil. Soc.* **34**, 429.
KENNEDY, SEABORG and SEGRÈ, 1939. *Phys. Rev.* **56**, 1095.
KIKUCHI, HUSIMI and AOKI, 1936. *Proc. Phys.-Math. Soc. Japan*, **18**, 188.
KIKUCHI *et al.* 1939. *Proc. Phys.-Math. Soc. Japan*, **21**, 381.
LANGER, MITCHELL and MCDANIEL, 1939 a. *Phys. Rev.* **56**, 380.
LANGER, MITCHELL and MCDANIEL, 1939 b. *Phys. Rev.* **56**, 422, 962.
LANGSDORF and SEGRÈ, 1940. *Phys. Rev.* **57**, 105.
LAWSON and CORK, 1940. *Phys. Rev.* **57**, 982.
LEA, 1935. *Proc. Roy. Soc.* A, **150**, 637.
LEWIS and BOWDEN, 1934. *Proc. Roy. Soc.* A, **145**, 235.
LIVINGOOD and SEABORG, 1939. *Phys. Rev.* **55**, 457.
MACLEAN, BECKER, FOWLER and LAURITSEN, 1939. *Phys. Rev.* **55**, 796.
MAY, 1936. *Rep. Progr. Phys.* **3**, 89.
MEITNER, 1925. *Z. Phys.* **34**, 807.
MITCHELL and LANGER, 1938. *Phys. Rev.* **53**, 505.
MITCHELL, LANGER and MCDANIEL, 1940. *Phys. Rev.* **57**, 1107.
MOON, 1937. *Rep. Progr. Phys.* **4**, 198.
MOON and TILLMAN, 1936. *Proc. Roy. Soc.* A, **153**, 476.
NELSON and HEBB, 1940. *Phys. Rev.* **58**, 486.
NORLING, 1938. *Z. Phys.* **111**, 158.
NORLING, 1939. *Naturwissenschaften*, **27**, 432.
OPPENHEIMER and SCHWINGER, 1939. *Phys. Rev.* **56**, 1066.
PEREY, 1939. *C.R. Acad. Sci., Paris*, **208**, 97.
PEREY and LECOIN, 1939. *J. Phys. Radium*, **10**, 439.
PLAIN, HERB, HUDSON and WARREN, 1940. *Phys. Rev.* **57**, 187.
PONTECORVO, 1936. *J. Phys. Radium*, **7**, 511.
PONTECORVO, 1938 a. *Phys. Rev.* **54**, 542.
PONTECORVO, 1938 b. *C.R. Acad. Sci., Paris*, **207**, 230.
PONTECORVO, 1939. *Nature, Lond.*, **144**, 212.
PONTECORVO and LAZARD, 1939. *C.R. Acad. Sci., Paris*, **208**, 99.
POWELL, MAY, CHADWICK and PICKAVANCE, 1940. *Nature, Lond.*, **145**, 893
RASETTI, 1935. *Z. Phys.* **97**, 64.
RICHARDSON, 1938. *Phys. Rev.* **53**, 124.
RICHARDSON, 1939. *Phys. Rev.* **55**, 609.

RICHARDSON and KURIE, 1936. *Phys. Rev.* **49**, 209 A; **50**, 999.
ROBERTS, 1940. *Phys. Rev.* **57**, 564 A, 1069 A.
ROSENBLUM, 1929. *C.R. Acad. Sci., Paris*, **188**, 1401.
ROSENBLUM, GUILLOT and PEREY, 1936. *C.R. Acad. Sci., Paris*, **202**, 1274.
ROTBLAT, 1939. *Nature, Lond.*, **144**, 248.
ROUSSINOW and YUSEPHOVITCH, 1939. *Phys. Rev.* **55**, 979.
RUTHERFORD and CHADWICK, 1929. *Proc. Camb. Phil. Soc.* **25**, 186.
RUTHERFORD, CHADWICK and ELLIS, 1930. *Radiations from radioactive substances*, chap. 12.
RUTHERFORD, LEWIS and BOWDEN, 1933. *Proc. Roy. Soc.* A, **142**, 347.
SACHS, 1940. *Phys. Rev.* **57**, 159, 194.
SAGANE, KOJIMA and MIYAMOTO, 1939. *Proc. Phys.-Math. Soc. Japan*, **21**, 728.
SAHA, 1939. *Naturwissenschaften*, **27**, 786.
SARGENT, 1939. *Canad. J. Res.* **17**, 82.
SAVEL, 1935. *Ann. Phys., Paris*, **4**, 88.
SCHNETZLER, 1935. *Z. Phys.* **95**, 302.
SEABORG and SEGRÈ, 1939. *Phys. Rev.* **55**, 808.
SEABORG, LIVINGOOD and KENNEDY, 1940. *Phys. Rev.* **57**, 363.
SEGRÈ, HALFORD and SEABORG, 1939. *Phys. Rev.* **55**, 321.
SIDAY, 1939. *Nature, Lond.*, **143**, 688.
SNELL, 1937. *Phys. Rev.* **52**, 1007.
SODDY, 1932. *Nature, Lond.*, **130**, 364.
STEWART, 1939. *Phys. Rev.* **56**, 629.
TILLMAN and MOON, 1935. *Nature, Lond.*, **136**, 66.
VALLEY and MCCREARY, 1939. *Phys. Rev.* **56**, 863.
DE VAULT and LIBBY, 1939. *Phys. Rev.* **55**, 322.
WALDMAN and COLLINS, 1940. *Phys. Rev.* **57**, 338.
WALKE, 1939. *Rep. Progr. Phys.* **6**, 16, §§ 5 and 6.
WALKE, 1940. *Phys. Rev.* **57**, 163.
WALKE, THOMPSON and HOLT, 1940. *Phys. Rev.* **57**, 171.
WARD, 1939. *Proc. Camb. Phil. Soc.* **35**, 322.
WARD, 1940 a. In course of publication.
WARD, 1940 b. Unpublished.
WATASE and ITOH, 1939. *Proc. Phys.-Math. Soc. Japan*, **21**, 389.
v. WEIZSÄCKER, 1936. *Naturwissenschaften*, **24**, 813.

THE BOHR THEORY OF NUCLEAR REACTIONS

By R. PEIERLS,
The University, Birmingham

§1. INTRODUCTION

THE theory of atomic nuclei has made progress along two, more or less distinct, lines. The first consists in describing the features of nuclei and their fate in collisions in terms of a small number of parameters, thus correlating the results of different observations. The second way is to try to determine all constants from first principles, starting from the elementary laws governing the interaction of nuclear particles.

This second aspect of the theory is as yet in a very primitive stage of development with the exception of some results concerning the very lightest nuclei (Peierls, 1940). The reason is that in the nucleus each particle is always in strong interaction with many others and every mechanical problem must therefore be treated as one involving many particles. The mathematical difficulties in the way of tackling such a many-body problem are formidable. It was the discovery of Bohr (1936) that, just because of this strong coupling between different particles inside a nucleus, the first aspect of the theory, the description of the behaviour of nuclei in general, is made very much easier. Since then, the phenomenological theory of nuclear reactions has made rapid progress (Bethe, 1937).

The present report is almost entirely concerned with this phenomenological aspect of the theory. We shall partly use very general arguments involving only the principles of wave mechanics, and partly conclusions drawn from the known properties of nuclei. We shall not attempt to distinguish the two types of reasoning rigorously, but in no case will it be necessary to use any specific assumption about the nature of the intranuclear forces.

§2. THE NOTION OF THE "COMPOUND NUCLEUS"

It is an obvious remark to say that in treating the collision of a particle (which we shall in general refer to as the "projectile") with a nucleus, one should consider both together as one mechanical system. This system is then in a state of positive energy, that is to say its energy is higher than that necessary to remove one of its constituents (e.g. the projectile) to infinite distance. The importance of treating all the particles concerned as one "compound nucleus" lies, however, in the fact that they will remain together for a time which is very long if measured on the nuclear scale, i.e. long compared to the time taken by a particle of a few Mv. to traverse a distance of a few times 10^{-13} cm.

This can be visualized, according to Bohr (1936), in the following way: when the projectile enters the nucleus it has much more energy than any of the other constituents, because it has come from a large distance and, unless it has lost some of its energy, will be able to escape again (see also Bohr and Kalckar, 1937). Actually, however, owing to the strong interaction inside the nucleus, it will at once lose most of this excess energy, which will be shared among many other particles, and hence none of the particles will have sufficient energy to overcome the attraction by the rest of the nucleus and to escape. Eventually, by a fluctuation, the whole excess energy may again become concentrated in one particle, which is then enabled to escape.

Because of this, the compound nucleus will stay together for some considerable time before breaking up, and during this time the possibility of escape can, for many purposes, be neglected. To this approximation the compound nucleus is then perfectly similar to an ordinary nucleus in a high state of excitation. This simple description makes it clear that the treatment applies to nuclei containing many particles, i.e. to all but the lightest nuclei.

Like an ordinary stable nucleus, the compound nucleus may be expected to have a spectrum of quantized energy levels. Since the nucleus is only approximately stable these levels are only approximately sharp. They will have a width Γ which, by the uncertainty principle, is related to their mean lifetime by the formula

$$\Gamma = \frac{\hbar}{\tau}, \qquad \ldots\ldots(1)$$

where $\hbar$ is $\frac{1}{2\pi}$ times Planck's constant. The fact that the lifetime is long on the nuclear scale means that the width is small compared to, say, a few Mv., but since the spacing of the levels decreases rapidly with increasing excitation this does not necessarily mean that it is less than the distance between neighbouring levels. We shall see that, in fact, the width usually exceeds the level distance in the case of compound nuclei created by bombardment with fast particles, but that in most nuclei it is less when the projectile is slow. In the latter case, the width of the levels is insufficient to bridge the gaps between them, and in such cases one will observe typical resonance phenomena. The probability of formation of a compound nucleus, and hence of producing a nuclear reaction (or capture or inelastic scattering), is only appreciable in the neighbourhood of these well-defined resonance levels. This conclusion is in good agreement with the observed resonance phenomena in the interaction of slow neutrons with nuclei.

The case in which the width of the levels is less than their distance—or, for brevity, the case of discrete levels—is the simplest from the point of view of the theory. In this case the impact of a projectile of a certain energy on the initial nucleus can only produce a compound nucleus in one well-defined quantum state. The properties of this quantum state are completely defined by its energy, and are, in particular, independent of the particular fragments which were

brought together by the collision. We can then predict the yield of all possible nuclear reactions if we know (*a*) the spectrum of the compound nucleus, i.e. the position of its resonance levels, and (*b*) the properties of each resonance level, i.e. its lifetime or width and the probabilities of its losing its excess energy in the various possible ways, e.g. by emission of a neutron, a proton, or an α-particle, or by radiation.

We cannot hope, of course, to predict the positions and properties of individual levels, but we can obtain some statistical results on the distribution of the levels and on the order of magnitude and variation with energy of the various disintegration possibilities.

The next section will deal with these problems, while §§ 4 and 5 are concerned with the connection between these data and the observed cross-sections for the different types of reactions.

In the case of overlapping levels, i.e. when the width of the levels exceeds their distance, there is for each energy more than one quantum state of the compound nucleus, and the properties of the system are therefore not uniquely determined by the energy, but may also depend on the way in which the compound nucleus has been formed. In this case the methods of the theory are less powerful than in the first case, but we shall discuss some results in § 6.

In the remaining sections we shall briefly review the various types of experiments that have been made, but an attempt at anything like a full review of the available material would exceed the scope of this report.

§ 3. THE LEVEL DISTRIBUTION AND THE STATE OF NUCLEAR MATTER

According to Bohr the strong interaction between the constituents of a nucleus has the effect that the excitation energy does not affect merely the motion of one particle in the nucleus, but that the excitation will be shared by a large number of degrees of freedom. It is then convenient to use the methods of statistical mechanics, which are appropriate to systems containing a large number of degrees of freedom.

If we denote by $Z(E)dE$ the number of states with energies between E and $(E+dE)$, the level density can be expressed in terms of thermodynamic quantities in the following manner (Bethe, 1936):

$$Z(E)=\left(\frac{d^2S}{dE^2}\right)^{1/2}e^S. \qquad \ldots\ldots(2)$$

Here S is the entropy. The units are so chosen as to make Boltzmann's constant equal to unity; in other words, the temperature scale is identified with the energy scale. This is convenient, since one will hardly ever be in a position to measure the temperature of a " hot " nucleus by means of an ordinary thermometer.

In order to obtain a reasonable picture of the properties of a nucleus, we must remember certain facts which follow from a discussion of the typical features of nuclear structure. It is known that the volume of a nucleus is approximately proportional to the number of particles it contains, and that the binding energy is,

roughly speaking, also proportional to it (Peierls, 1936). For very heavy nuclei this view is not strictly correct because of the effect of the electric charge, but it seems a reasonable idealization to think of this effect as simply an addition to the energy of the nucleus, which does not vary much with its internal state of motion.

In this idealized picture a nucleus may be compared with ordinary macroscopic substances like solids, or liquids, of uniform density, in which also the energy content, the mass, etc., are proportional to the volume. On a nuclear scale, the "temperature" obtained in all practical cases is very low and the physical state of the matter forming the nucleus would, on this account, be expected to be similar to that of ordinary substances at very low temperatures, i.e. to the solid state. There is, however, one difference. In an ordinary solid the atoms form a regular crystal lattice, and each atom is at or near its position of equilibrium within this lattice. Owing to quantum effects the positions of the atoms show a certain amount of fluctuation about the equilibrium positions (zero-point energy) which is closely connected with the uncertainty principle. In consequence of the zero-point motion the particle will stray from its equilibrium position to distances of the order of

$$(\hbar/2m\omega)^{1/2}, \qquad \ldots\ldots(3)$$

where m is its mass and ω the frequency of vibration about the equilibrium position. In the nucleus the mass of the particle is of the same order of magnitude as the mass of the ions making up the lattice of an ordinary crystal; the forces, and hence the vibration frequencies, are much larger, and hence the amplitude of the zero-point motion is less, but this reduction is not nearly in the ratio of the linear dimensions. In a nucleus the amplitude of the zero-point motion is comparable with the distances between the particles, and this means that no regular configuration can be maintained (Bohr and Kalckar, 1937).

This leads one to think that the state of matter in a nucleus should with more justification be compared to that of a liquid. This analogy, too, has its limitations, since the reason for the breakdown of an ordered crystalline arrangement is different. In an ordinary liquid the molecules are not thrown out of their equilibrium positions by zero-point motion, but by their heat motion, and because of this, an ordinary liquid necessarily has an appreciable entropy.

Probably a closer analogy is provided by the exceptional behaviour of liquid helium, which remains a liquid at the absolute zero. The cause of this behaviour is the zero-point motion, important because of the small mass and weak attractive forces.

For purposes of comparison one sometimes uses the extreme limiting case of weak forces, in which one considers the nucleus simply as consisting of a number of particles moving independently within a certain volume with no interactions between them. The properties are then the same as those of a gas, more accurately of a degenerate Fermi gas, like the electrons in a metal.

The energy content of such a substance at low temperatures is then

$$E = B \,.\, T^n, \qquad \ldots\ldots(4)$$

where T is the temperature, and B and n are constants. The exponent n is 4 for a solid, and probably also for a " quantum liquid " of the helium type, and 2 for a Fermi gas.

Using then the usual thermodynamic relations between energy and entropy, we have for the entropy S, i.e. for the exponent of (2):

$$S = \frac{n}{n-1} B^{1/n} E^{1-1/n}. \qquad \ldots\ldots(5)$$

This becomes $4/3\ B^{1/4}E^{3/4}$ for $n=4$, and $2\ (BE)^{1/2}$ for $n=2$. The order of magnitude of the constant B follows from the fact that the lowest excited level of a heavy nucleus is of the order of a few hundred kv., and that there, although the statistical formula (2) is no longer applicable, the entropy must be of the order of magnitude unity. It follows from this that at an excitation energy of about 6 Mv. (the energy of a compound nucleus formed by adding a slow neutron or proton to a nucleus in its normal state) the exponent of (2) is large; hence the level density is high and increases rapidly with increasing energy.

The dependence of level density on the mass number may, in the limiting case of a very large nucleus, be obtained from the fact that the constant B in equation (4) is proportional to the mass (or volume) provided the body is large enough for surface effects to be negligible. This, however, is a very crude approximation, since the diameter of the heaviest known nuclei contains only about 7 particles, and hence more than half the particles lie on the surface.

The formula (2) is very sensitive to the value of the exponent, and as we can at best obtain only a rough estimate for the constants it is impossible to calculate the level density. Experimentally, the level density can be estimated from a statistical analysis of the resonance levels observed in the collisions of slow neutrons with nuclei, and it is found that, at an excitation energy of about 6 Mv., the distance between the levels is about 100 volts or less (Bethe, 1937). The values which one obtains then for B in (5) with $n=2$ or 4 are quite compatible with the requirement that the first excited levels, i.e. the energies for which S becomes comparable with unity, are of the order of a few 100 kv.

If we accept these figures we can calculate the " temperature " of the compound nucleus which, both for $n=2$ and $n=4$, becomes 0·6 Mv. This is one-tenth of the energy content, and hence, of the degrees of freedom of the nucleus, about one-tenth are excited in this range. This shows, on the one hand, that it is reasonable to assume the excitation energy spread over many levels as required for the Bohr theory, and, on the other hand, that only a small fraction of the total number of degrees of freedom are excited and that we are therefore justified in using the equation (4) applicable to a " cold " substance.

§4. DISINTEGRATION

The states of the compound nucleus which we have considered in the preceding section are unstable. The nucleus will, in the course of time, get rid of the excess energy in some way. The most important kinds of transition that may

take place are:—(*a*) the emission of a particle (neutron, proton, α-particle), (*b*) emission of radiation (γ-rays), (*c*) emission of a positive or negative electron (β-rays) and (*d*) fission of the nucleus.

Because of the effect of the potential barrier, the emission of neutrons is generally much more probable than that of charged particles except for light nuclei or extremely high energies. β-decay is, in general, so slow that it cannot compete with the other processes except in the case of states of exceptionally long life. In general, β-decay from highly excited states is negligible. Fission is restricted to certain very heavy nuclei and will be considered in § 10. Hence the most important effects which determine the lifetime of an excited state of the compound nucleus are the emission of a neutron and the emission of radiation.

Since the possibility of decay is responsible for the width of the energy level by equation (1) it is convenient to attribute to each process a " partial " width, which is simply $\hbar$ times the probability, per unit time, of this process. The total width is the sum of the partial widths for all possible transitions.

The partial width Γ_N of the compound state for neutron emission depends strongly on the energy of excitation as compared with the energy required to break off a neutron. If the excitation energy is less than this binding energy the probability is evidently zero. If the excitation energy only just exceeds the threshold, the neutron must emerge with very low kinetic energy.

In this case the neutron can only escape with zero angular momentum (an angular momentum of one quantum unit would mean that the line of motion of the neutron did not pass through the centre of the nucleus, but passed at a distance equal to the neutron wave-length; this is extremely improbable if the neutron wave-length exceeds the size of the nucleus). From the conservation of angular momentum, it follows then that the escape is possible only for certain of the states of the compound nucleus.

For these states it can be shown (Bethe and Placzek, 1937) that the probability of neutron escape (and hence the partial width due to this effect) is proportional to the velocity of the escaping neutron. As the energy of excitation, and hence also the energy of the escaping neutron, increases further, the partial width increases for two reasons: one is that the neutron velocity increases, and the second is that the energy is now sufficient to produce a free neutron plus an excited residual nucleus. One is dealing, therefore, not with a single transition, but with many, the number increasing rapidly as the increasing energy makes more excited states of the residual nucleus available.

One may also ask for the partial width Γ_{N0} due to the escape of a neutron with the full available energy, the residual nucleus being in the ground state. This is, for energies near the threshold, identical with the partial neutron width Γ_N as defined before, since the energy is then insufficient to excite the residual nucleus. Hence initially Γ_{N0} increases proportionally to the neutron velocity. However, it reaches a maximum and decreases again at higher energies, and therefore at high energies represents only a small fraction of the neutron width Γ_N.

This can be shown by applying the law of detailed balancing, one of the fundamental laws of statistical mechanics, to the escape of a neutron from the compound nucleus and to the inverse process, viz. the formation of the compound nucleus by collision of a neutron with the residual nucleus in its ground state (Landau, 1937 ; Weisskopf, 1937 ; Frenkel, 1936 ; Kalckar, Oppenheimer and Serber, 1937).

This leads to the relation *

$$\Gamma_{N0}Z(E)=\frac{2\sigma_N}{(2l+1)\lambda^2}, \qquad \ldots\ldots(6)$$

which refers to a compound state of a given angular momentum, and the emission of a neutron with a given angular momentum l.

λ is the de Broglie wave-length of the escaping neutron, and σ_N is the effective cross-section for the formation of the compound nucleus in a collision of a neutron of angular momentum l with the residual nucleus. $Z(E)$ is the level of density as before. Now we shall see in § 6 that the cross-section cannot exceed the value $(2l+1)\lambda^2/\pi$ and that it is likely that it becomes just equal to this value at high energies. Hence Γ_{N0} must always remain less than

$$\frac{2}{\pi}\frac{1}{Z(E)}, \qquad \ldots\ldots(7)$$

which is of the order of magnitude of the distance between neighbouring energy levels, and most probably it is equal to this limit for high energies. Since we know that the level distance decreases rapidly with increasing energy the same must ultimately hold for Γ_{N0}.

Thus we see that at high energies Γ_{N0} is only a small fraction of Γ_N, and hence, the re-emission of the neutron with the full available energy is very unlikely. The process is in many ways analogous to the evaporation of molecules from a liquid drop. In this case the evaporating molecule is very unlikely to take away the whole heat content of the drop, but its mean energy will only be of the order of kT. A quantitative study of this analogy was made by Frenkel (1936), but it was pointed out by Weisskopf (1937) that the analogy is limited by the fact that the binding energy of the neutron is comparable to the excitation energy, and that therefore the temperature of the residual nucleus is appreciably lower than that of the compound state. This somewhat complicates the energy distribution of the escaping neutrons, but it is still roughly of the type of a Maxwell-Boltzmann distribution.

As for the partial width due to the emission of radiation, this is not particularly small when the energy is in the neighbourhood of the threshold for neutron emission or even below it, and it is therefore practically equal to the total width of these states. Again, this process may either lead directly to the ground state of the compound nucleus, or to an intermediate state of lower excitation. We have seen, however, that near the threshold energy the level density is already very

* For the sake of simplicity we neglect the spin of the projectile and that of the initial nucleus in all formulae of the text. The corrections for this omission were given by Bethe (1935 and 1937).

large and that there are a large number of excited levels with energies between the neutron threshold and the ground state.

Transition to these levels gives rise to radiation of lower frequency than that to the ground state, and hence the transition probability per state is somewhat less, but this is compensated for by the large number of states, and most of the radiation will consist of quanta containing only a fraction of the excitation energy. We may again liken the process to the emission of radiation by a black body which consists mainly of quanta of the energy kT; but the formula obtained in this way for the spectral distribution does not hold for transitions to states of low excitation, since in that region the nucleus is no longer a black body.

§ 5. NUCLEAR REACTIONS: THE DISPERSION FORMULA

The knowledge of the position of the energy levels of the compound nucleus and their partial widths is essential and, in many cases, sufficient for predicting the reactions that will take place in a given collision.

We shall restrict our discussion in this section to the case of discrete levels, i.e. that in which the total width of each level is less than its distance from the neighbouring levels of the same total angular momentum. Then, once a compound nucleus is formed, its probability, per unit time, of disintegrating by means of a certain process b, is $1/\hbar$ times the partial width Γ_b due to this process. Since the mean lifetime of the compound state is $\Gamma/\hbar$ where Γ is its total width, the probability of the process b is

$$\Gamma_b/\Gamma. \qquad \ldots\ldots(8)$$

In order to obtain the effective cross-section for the process b to occur in a collision, (8) must be multiplied by the effective cross-section σ_c for the formation of a compound nucleus:—

$$\sigma_b = \sigma_c \Gamma_b/\Gamma. \qquad \ldots\ldots(9)$$

In the neighbourhood of a resonance level, i.e. of a level of the compound nucleus, the cross-section for the formation of a compound nucleus is given by Breit and Wigner (1936) as:

$$\sigma_c = \frac{\lambda^2}{\pi}(2l+1)\frac{\Gamma_a\Gamma}{\Gamma^2+4(\Delta E)^2}. \qquad \ldots\ldots(10)$$

Here ΔE is the difference between the actual energy and the position of the resonance level. l is the angular momentum of the incoming particle, and it has been assumed, for simplicity, that the original nucleus and the projectile have no spin. Γ_a is the partial width of the resonance level due to the process a, which consists of the re-escape of the original projectile, leaving the nucleus in its initial state.

This equation can be made plausible in the following way: if we were dealing with a large number of projectiles of slightly different energies covering an energy interval E_1 to E_2 which includes the resonance level, we could use statistical arguments, and, from detailed balancing, obtain equation (6), in which for the number of levels per unit energy we must insert $1/(E_1-E_2)$, since there is just

one level in this interval, for the partial width for re-escape Γ_a instead of Γ'_{N0}. The mean cross-section for the energy interval will therefore be

$$\frac{\lambda^2}{2}(2l+1)\frac{\Gamma_a}{E_1-E_2}. \qquad \ldots\ldots(11)$$

It can easily be verified that this is indeed the average of (10) over an energy interval (E_1, E_2) which is large compared with the width Γ.

This justifies (10) in the mean. As to the actual dependence on E, equation (10) is analogous to the resonance curve for damped harmonic oscillations, and this shape of the resonance curve must be obtained in all cases in which the decay of the excited state is exponential; in other words, when the disintegration probability of the compound nucleus does not vary during its lifetime. The situation is in this respect completely similar to the natural width of spectral lines. Hence the partial cross-section for the process b as a result of the collision a is

$$\frac{\lambda^2}{\pi}(2l+1)\frac{\Gamma_a\Gamma_b}{\Gamma^2+4(\Delta E)^2}. \qquad \ldots\ldots(12)$$

This cross-section, if it is divided by the square of the wave-length of the initial particle, is symmetrical in a and b, as required by the law of detailed balancing.

Some care has to be exercised, however, in regard to the definition of the partial widths Γ_a, Γ_b in (12). They do not depend only on the properties of the resonance level to which they belong, but also on the energy of the projectile in relation to the resonance energy. In other words, they may, in the extreme wings of the resonance curve, have different values from those in the centre.

This can be made obvious by considering the following example: Assume the initial process, a, to be the excitation of a nucleus by radiation, and the result, process b, to be the emission of a slow neutron from a certain residual nucleus. Assume, further, that the energy of the escaping neutron is, for exact resonance, comparable to the width of the resonance level. In this case if we have exact resonance, there will be an appreciable chance of the neutron escaping. On the other hand, if we irradiate with radiation of a frequency rather below the resonance frequency, we know from energy conservation that the compound nucleus which we produce must have an energy somewhat below the normal, and that the slow neutron must come out with correspondingly lower energy. If we take an energy sufficiently far from the centre of the resonance curve (12), the neutron (if emitted) would have to have negative energy, and hence the probability of its emission is zero. Similarly, if the neutron has to come out with very low energy, its emission probability is low, varying proportionally to the neutron velocity, and we have to insert a correspondingly lower partial width in (12).

The same must obviously happen if we consider the inverse process, i.e. the bombardment of a nucleus by a very slow projectile, a case which is of greater practical interest.

Our example shows, also, that this variation of the partial width over the resonance level is not relevant except in cases where the width of the resonance

level is comparable to the kinetic energy of the projectile or of one of the possible emitted rays, or when we want to study the extreme portions of the " wings " of a resonance peak. The main cases in practice are that in which the projectile is a thermal neutron, and that of fairly light nuclei.

Equation (12) applies only if there is not more than one resonance level in the neighbourhood of the energy under consideration. In the more general case it has to be replaced by a sum

$$\sigma_{ab} = \frac{\lambda^2}{\pi}(2l+1)\left|\sum_n \frac{u_{na}u_{nb}}{2(E-E_n)-i\Gamma_n}\right|^2, \qquad \ldots\ldots(13)$$

where each term refers to a particular resonance level, and where u_{na}, u_{nb} are quantities characteristic of the resonance level n and the process a or b respectively, such that

$$u_{na}^2 = \Gamma_{na} \quad \text{and} \quad u_{nb}^2 = \Gamma_{nb}. \qquad \ldots\ldots(14)$$

These u may be positive or negative. Equation (13) is quite analogous to the formula for the coherent or fluorescent scattering of light by atoms. This has been most extensively studied in its influence on dispersion, and hence (13) is generally known as the " dispersion formula ".

The one-level formula (12) was first derived by Breit and Wigner (1936) and generalized to the many-level formula (13) by Bethe and Placzek (1937), who also gave the more general result for the case of the initial nucleus and the projectile possessing a spin. A more rigorous derivation was given by Kapur and Peierls (1938), who showed that the formula is restricted to the case in which the width of the levels is less than their distance apart, and takes a more complicated form when the levels overlap. (Cf. also Breit, 1940; Siegert, 1939.)

In writing (13) it has been assumed that the process b is not simply the re-escape of the initial particle; in other words that we are not dealing with elastic scattering, for which $a = b$. In the latter case the formula has to be supplemented by a further term, so that it reads

$$\frac{\lambda^2}{\pi}(2l+1)\left|\sum_n \frac{\Gamma_{na}}{2(E-E_n)-i\Gamma_n} + S_l\right|^2. \qquad \ldots\ldots(13\,a)$$

Here S_l is of the order of magnitude of the amplitude of the wave scattered by a hard elastic sphere of the same radius as the nucleus in the case of a neutron, and of the wave scattered by a charged sphere in the case of a charged particle. This term is generally referred to as the " potential " scattering, since it represents the scattering merely by virtue of the interaction potential between the particle and the nucleus as a whole, without a compound nucleus being formed.

For charged particles this potential scattering represents the main contribution unless their energy is high enough to overcome the electrostatic repulsion which tends to prevent the formation of a compound nucleus as well as the re-escape of the projectile, if the binding energy, i.e. the work necessary to remove the particle to infinite distance, is given.

For neutrons, the contribution from potential scattering is of the order of magnitude of the geometrical cross-section of the nucleus and therefore it is, for

slow neutrons, negligible in comparison with the other terms of (13 *a*) which are in the neighbourhood of each resonance level, of the order of magnitude of the square of the wave-length, except when one considers a very rare process for which Γ_b is very small. The latter case arises probably in the scattering of slow neutrons by heavy nuclei, in which the re-escape of the neutron is a much less frequent event than the emission of the excess energy in the form of radiation. Hence the contribution of the sum in the dispersion formula (13 *a*) is very small, and the main contribution comes from the slowly varying term, S_l, so that the scattering of slow neutrons by heavy nuclei does not show any reasonance effects.

The discussion of this " potential scattering " is somewhat complicated by the presence of the distant terms in the sum (13 *a*). Near a resonance level one term in this sum predominates, and the behaviour in the neighbourhood of such a resonance level is then substantially as given by the " one-level " formula (12).

All other terms of the sum together give a contribution which varies slowly over the resonance level and merely produces a general background. It is very difficult to estimate this background, since the sum may converge slowly. Indeed, the density of energy levels increases with increasing energy, and the sum would diverge if the numerator did not become very small for the high levels. Hence it is clear that the discussion of the general background, or, in other words, the magnitude of the cross-section between the resonance levels, is impossible without knowing something of the properties of resonance levels of very high excitation energy.

In the formula for the cross-section of any process other than elastic scattering, the background is given entirely by the distant terms of the sum (13); in the case of elastic scattering it is due to both these and the " potential scattering ". In fact, the distinction between the two contributions is not quite unambiguous. The mathematical treatment of the compound nucleus involves an assumption as to the nuclear radius which, to a certain extent, is arbitrary. One has to count any particle closer than a certain radius as part of the compound nucleus, whereas any particle beyond this counts as free. If one alters one's choice of this radius, one obtains different results for the properties of high resonance levels, but also a different " potential scattering ", so that the physical results remain unchanged.

One may conjecture that, if a suitable size for the nucleus is chosen and used in the calculations, the infinite sum in the dispersion formula (13) will converge rapidly. This would mean in the case of inelastic processes that the background between the levels should be mainly due to the contributions from nearby levels, and in the case of elastic scattering mainly to the " potential scattering ". No proof of the truth of this conjecture, however, has yet been given.

§6. THE CASE OF OVERLAPPING LEVELS

With increasing excitation energy the width of the levels increases, because a highly excited nucleus has more possibilities of breaking up, and their distance decreases as explained in §3. Hence we eventually reach a region in which the

level distances become comparable, so that the resonance curves due to neighbouring levels merge into each other. At still higher energies the width will be large compared to the level spacing, and then the excitation curves of all reactions will have lost all their resonance features and the cross-sections will vary regularly with energy.

In discussing the magnitudes of the cross-sections we must remember, as before, that the observed cross-section is the sum of the contributions due to projectiles with all possible values of the orbital angular momentum l (in units $h/2\pi$). The particles with a certain l in the incident beam just fill a cylinder of cross-section $(2l+1)\lambda^2/\pi$, and since not more than all particles can produce a collision, the maximum possible cross-section for any type of collision for particles of angular momentum l is

$$(2l+1)\lambda^2/\pi.$$

We have already seen that if the wave-length is much greater than the nuclear radius, only particles with $l=0$ will contribute: in this case the maximum cross-section is

$$q=\lambda^2/\pi. \qquad \text{.....(15)}$$

If the wave-length λ is small compared to the nuclear radius R, one knows that particles with angular moments up to $2\pi R/\lambda = l_R$ have a chance of colliding with the nucleus, and hence the total cross-section may, at most, equal the sum of their maximum contributions, i.e.

$$q=2\sum_0^{l_R}(2l+1)\lambda^2/\pi=4\pi R^2. \qquad \text{.....(16)}$$

One may combine (15) and (16) by defining a maximum cross-section

$$q=4\pi(R+\lambda/2\pi)^2, \qquad \text{.....(17)}$$

which will represent an approximate upper limit also in the intermediate case, in which λ and R are comparable.

One can further reduce this limit in the case of inelastic collisions. According to a general theorem of wave mechanics (Bohr, Peierls and Placzek, 1941), the total cross-section for all collisions, except those leading to elastic scattering, cannot exceed one-quarter of the maximum, and, moreover, if it reaches this value, there must also be an elastic cross-section of at least another quarter.

The first half of this statement is eminently plausible in the case of very short waves, because then it means a limit of πR^2 for the inelastic cross-section. Now for short waves the approach of the projectile to the nucleus can be described by classical mechanics, and it is reasonable in that case to find a limit equal to the geometrical cross-section of the nucleus. The elastic scattering of the same amount, which we have to expect in addition, becomes more plausible if we remember that it is small-angle scattering, composed of particles which have not been deflected by an angle of more than about $\lambda/2aR$ from the incident beam. In the case of macroscopic bodies, for which λ/R is extremely small, this cannot be distinguished from the incident beam. Actually, this small deflection is due to

diffraction taking place at the edge of the " shadow " thrown by the scattering body.

For the scattering of fast particles by heavy nuclei, on the other hand, the ratio between λ and R is never so small as to make this effect difficult to observe, and measurements on this diffraction pattern would lead to useful conclusions as to the size and shape of the nucleus.

If we disregard this " shadow " effect, it is reasonable to expect that the total remaining cross-section will be just equal to πR^2 in the limiting case of very short waves. In this case, in which classical mechanics is applicable to the approach, we can predict that the cross-section comprising all those particles whose path hits the nucleus is *equal* to πR^2. Any particles outside this section cannot possibly hit the nucleus, and, on the other hand, it seems impossible for a particle to go right through the nucleus without suffering deflection or energy loss. (This conclusion may not apply to very fast projectiles, like those present in cosmic rays, which have energies far exceeding the interaction energies between particles inside the nucleus.)

Hence if we denote the ratio of the total cross-section (exclusive of the " shadow ") to its maximum possible value, q, by ξ, ξ represents a " sticking probability ", i.e. the chance of *something* happening to a particle which strikes the nucleus. Then $\xi = 1$ for fast neutrons, whereas for slow neutrons ξ is proportional to the neutron velocity, according to the results of the preceding section.

In the case of charged particles, it is convenient to allow for the scattering by the electrostatic potential in defining the sticking probability. In other words one considers as collisions only those cases in which a compound nucleus has actually been formed, and expresses the probability of this happening in terms of the number of particles actually reaching the nucleus. If that is done, the sticking probability may be expected to behave in a very similar way to the case of neutrons, in particular to reach unity at energies of a few Mv.

The actual rise of the sticking probability from zero at low energies to unity can only be obtained from a more detailed theory of the nucleus. An approximate theory of this problem has been given by Bethe (1940), who considers the nucleus as a potential field, allowing for the possibility of inelastic collisions by introducing an absorption coefficient, which represents, schematically, the particles which, by inelastic collisions or with other constituents of the nucleus, lose some of their energy and therefore cease to form part of the original coherent beam. The results one obtains in this way depend essentially on the gradient of the absorption coefficient in the neighbourhood of the surface of the nucleus. If it is assumed that the rise of the absorption coefficient from zero to large values takes place in a layer of which the depth is about 10^{-13} cm., one obtains reasonable results for the variation of the sticking probability with energy.

According to equation (6), which was derived from very general statistical considerations, the sticking probability is equal to

$$\xi = \frac{\pi}{2} \Gamma_c \, Z(E), \qquad \ldots\ldots(18)$$

where Γ_a is the partial width for re-emission of the projectile. Hence, this partial width cannot exceed the value $\frac{2}{\pi}\frac{1}{Z(E)}$, which is about the mean distance between the levels, and it must become equal to it for high energies. Hence, when the sticking probability has reached the value unity, the partial width for a particular kind of breaking up is comparable to the level distance. Since at these energies the compound nucleus can break up in many ways, the total width is then necessarily larger than the distance, i.e. we are dealing with the case of overlapping levels.

The above results refer to the magnitude of the total cross-section. One wants also to know what part of this cross-section is due to a particular process. This is again answered by equation (9), where the Γ represent partial widths, or, apart from a factor $\hbar$, probabilities per unit time. However, these probabilities may now depend on the way in which the compound nucleus has been formed (e.g. by collision of a fast neutron with a normal nucleus of mass number $(A-1)$ and atomic number Z, or of a slow neutron with an excited nucleus, or a proton with the nucleus $(A-1)$, $(Z-1)$). This difference from the case of discrete levels has its origin in the fact that, in the case of discrete levels, only one of the states of the compound nucleus, if any, satisfies the resonance condition to within its width. Hence the nucleus is formed with overwhelming probability in this quantum state, and, being in a definite quantum state, all its properties, including the disintegration probabilities, are fixed.

On the other hand, if the level width exceeds the distance, the resonance condition is always satisfied for several quantum states, and hence the compound nucleus may be formed in any of these states (or indeed in any combination of them). Hence its properties may show a certain amount of variation according to the way in which it has been formed. It is unlikely that this variation will in practice be very serious, except that it will always increase the probability of re-emission of the projectile leaving the original nucleus. This is necessary to allow for the " shadow " phenomenon which we discussed above.

§ 7. COLLISIONS OF SLOW NEUTRONS WITH NUCLEI

The term " slow " neutron generally refers to neutrons with energies below about 1000 volts, amongst which neutrons with thermal energies are of particular interest. At these energies the escape of a neutron is a rare event, because of the small kinetic energy, and the life of the compound nucleus is practically limited by the emissions of radiation (" radiative capture ") except in the case of the very lightest nuclei. This follows most directly from the fact that resonance capture of neutrons has been observed, but not resonance scattering.

The energies of some of the resonance levels have been observed by measuring absorption curves with different detectors. A number of elements have resonance levels of energies less than, say, 100 volts, and this fact leads one to the rough estimate of the level density which we used in § 3.

As regards the shape of the resonance curve in the neighbourhood of each level, it has not so far been possible to observe this curve in full detail, because

of the difficulty of obtaining or observing neutrons of homogeneous energy. One can, however, deduce from some experiments two important quantities: (*a*) the integral over the resonance peak, and (*b*) the cross-section at maximum resonance. The first determines the number of neutrons absorbed into this level from a beam in which the neutron energies are uniformly spread over a certain energy range ; the other determines the absorption coefficient of the substance if the same substance is used as an indicator, when the influence of neutrons with energies far from the resonance level in question is eliminated in some way.

The integrated cross-section is, according to equation (11), proportional to the partial width for re-emission, Γ_a, which, according to the results of § 5, is proportional to the neutron velocity:

$$\Gamma_{N0} = \sqrt{(E . E_1)}. \qquad \ldots\ldots(19)$$

By using the experimental figures for the integrated width (Hornbostel, Goldsmith and Manley, 1940; Manley, Goldsmith and Schwinger, 1939) it is now possible to estimate the factor of proportionality, E_1, which is of the order of magnitude of 3×10^{-6} volt.

In considering the cross-section at resonance, one must ensure that the resonance curve is not further broadened by the Doppler effect produced by the thermal motion of the initial nucleus. This Doppler width is greater for the lighter nuclei, and for resonance levels of higher energy.* The Doppler width can, of course, be easily calculated, and one has to make sure in each case that the true width of the level as determined from the observations is greater than the Doppler width. If the Doppler width is negligible, we have from (10) for the cross-section at resonance,

$$\sigma = \frac{\lambda^2}{\pi} \frac{\Gamma_{N0}}{\Gamma},$$

where we have put $l = 0$, and inserted Γ_{N0} for Γ_a. If we know the integrated cross-section from the result under (*a*), we know Γ_a, and hence can deduce the width Γ of the level. This is in practice equal to the radiation width. From such observations the radiation width is found to be of the order of 0·3 volt (Hornbostel, Goldsmith and Manley, 1940 ; Manley, Goldsmith and Schwinger, 1939).

As pointed out before, this total radiation width is the resultant effect of a large number of possible transitions, of which the majority belong to the emission of only a fraction of the total excitation energy. These leave the nucleus in an excited state, from which further radiation will be emitted. Indeed, it has been found (Griffiths and Szilard, 1937) that the number of photons emitted by cadmium as a result of neutron capture is considerably larger than the number of absorbed neutrons.

In the particular case of thermal neutrons, one has to consider the dependence

* The effect of the Doppler broadening is the same as that of a constant increase in the velocity of the neutron, and this implies a higher energy increase if the velocity is high.

of the partial width on the neutron velocity ; in the thermal region the partial width Γ_a in each term of (13) is proportional to the velocity, and since the wave-length is inversely proportional to v, the cross-section is proportional to $1/v$. The factor of proportionality depends on those terms of (13) which belong to energy levels in the neighbourhood of zero. A very large capture cross-section will, in particular, be obtained if one level happens to be very close to zero.

With increasing neutron energy the capture cross-section continues proportional to v as long as the denominators of (13) are practically constant, i.e., as long as the neutron energy is small compared either to the energy or to the width of the nearest resonance level. In the case of the collision of neutrons with such light nuclei as B or Li, the most probable process is not the emission of radiation but the emission of an α-particle. The kinetic energy available for the latter is very high and the potential barrier opposing its escape is low, since we are dealing with a light element. Hence the levels of the compound nucleus are here very wide, and, moreover, they are, in a light element, very widely spaced. Because of this, the $1/v$-law holds for these nuclei up to fairly high energies, at any rate throughout the region normally referred to as that of " slow " neutrons. Hence the absorption of neutrons in boron is often used as a standard in order to determine neutron energies (Frisch and Placzek, 1936).

§8. FAST NEUTRONS AND CHARGED PARTICLES

If the energy of the neutron is increased, the partial width due to neutron emission begins to compete with that due to radiation. If one extrapolates the equation (19) for the neutron width, this should reach the value of 1 volt at an energy of about 0·3 Mv., but actually Γ_N will rise more rapidly than this. On the other hand, the radiation width cannot be expected to change very much in this range since the natural zero level for the radiation problem is not the energy at which a neutron can just escape with no kinetic energy, but the ground level of the compound nucleus. While the distance from this ground level changes, say, from 7 to 7·5 Mv., we cannot expect any violent changes in the radiation properties.

Hence, at least above about 0·3 Mv., the emission of a neutron is more probable than that of radiation; in other words, the cross-section for scattering (as we saw in § 4, mostly inelastic scattering) is greater than that for capture.

Hence inelastic scattering is the main phenomenon for fast neutrons, and the cross-section for scattering is then practically equal to the total cross-section, which, as we have seen, is of the order of magnitude of the geometrical cross-section of the nucleus, when the wave-length is smaller than the nuclear radius.

As we saw in § 4, the energy of the neutron which is re-emitted from the compound nucleus is only of the order of the nuclear " temperature ", i.e. probably less than 1 Mv. The remainder of the kinetic energy of the incident neutron is left in the nucleus as excitation energy, and will eventually be emitted in the form of radiation. If the kinetic energy of the neutron is higher than the binding energy of a neutron in the nucleus, the residual nucleus may still be left

in a state of sufficiently high excitation for a second neutron to escape from it. This will be likely to happen if the energy width of the compound nucleus which is left after the first neutron has escaped is due to neutron escape rather than to radiation, i.e. if after the escape of the second neutron the nucleus still has an excitation energy of 1 Mv. or more.

The compound nucleus formed by the capture of a neutron, or the excited nucleus left after an inelastic collision, may, by a radiative transition, eventually reach a state of very long life (metastable state) in which the transformation by β-decay can successfully compete with the emission of further radiation. This leads to interesting phenomena connected with nuclear isomerism which lie, however, outside the scope of the present report (Goldhaber, Hill and Szilard, 1939).

Much theoretical work has also been done (Weisskopf and Ewing, 1930 ; Bethe and Konopinski, 1938) on the case in which charged particles are used as projectiles. In this case there is no analogy to the case of slow neutrons, since slow charged particles are prevented by the electrostatic repulsion from approaching sufficiently close to the nucleus to cause a reaction. In general, the phenomena caused by fast charged particles are very similar to those produced by fast neutrons, except that we should expect a neutron rather than the original charged projectile to be the first to emerge, again because the neutron is not delayed by the electrostatic potential barrier. This holds if, in the compound nucleus, protons and neutrons are about equally strongly bound, i.e. for most stable heavy nuclei. For some light nuclei, however, the ratio of charge to mass is high, and they require more energy to split off a neutron than a proton. Similarly, the emission of an α-particle may leave a more stable structure than that of a neutron or a proton. In such cases the emission of a charged particle may be a more frequent event than that of a neutron. In the same way light nuclei, when bombarded with neutrons, may sometimes emit charged particles rather than neutrons or radiation.

The deuteron has, as a projectile, an intermediate position between charged and neutral particles. Since it consists of a proton and a neutron which are only loosely held together, the reaction with a nucleus may take place in such a manner that the neutron enters the nucleus, whereas the proton stays outside and never has to overcome the full effect of the electric repulsion (Oppenheimer and Phillips, 1935 ; Bethe, 1938 ; Lifshitz, 1938).

§9. NUCLEAR PHOTO-EFFECT

The disintegration of nuclei by γ-rays, which has been extensively studied by Bothe and his collaborators (Bothe and Gentner, 1937), is in a sense the inverse of the radiative capture of neutrons. The difference is, however, that in the one case we are dealing with a neutron and a nucleus in its ground state, while the emitted photon will in general leave the nucleus in an excited state. On the other hand, in the photo-effect we start with a photon and a nucleus in its normal state, and end up with a neutron and (in general) an excited nucleus.

The wave-length of the γ-rays used for such experiments is greater than the

nuclear radius (for $h\nu = 15$ Mv., $\lambda = 10^{-10}$ cm.). Hence the maximum cross-section is λ^2/π. The ratio of the actual cross-section to this maximum is, according to (18),

$$\xi = \frac{\pi}{2}\Gamma_a \frac{1}{Z(E)}$$

where $Z(E)$ is the number of levels (with suitable angular momentum) per unit energy, and Γ_a is the partial width for re-emission of the " projectile " i.e. in our case for the re-emission of the full excitation energy in the form of radiation, leaving the nucleus in the ground state. This is much less than the partial width due to all radiative processes, since we have seen that in the majority of cases the photon emitted in the first instance has only an energy of the order of the nuclear " temperature ". Hence, although the radiation width is, at these energies, probably much in excess of the level spacing, the partial width for the transition to the ground state is much less, and hence the " sticking probability " is small.

From the experiments, the sticking probability is found to be about 10^{-3}, provided that the cross-section for the emission of neutrons is identified with the total cross-section. This is reasonable, since we have seen that at these high energies the partial width for neutron emission certainly exceeds that for radiation.

§ 10. NUCLEAR FISSION

It was found recently that some heavy nuclei could disintegrate in an entirely different way from the ones discussed above (Walke, 1940). These nuclei, when bombarded by neutrons, split into two fragments, of roughly equal size, which are in general highly excited, and therefore decay further by emitting neutrons; most of these fragments also show radioactivity. This process is known as "fission". Its compatibility with energy conservation is obvious if one considers the binding energy as a function of the nuclear mass. In the first part of the system of elements the binding energy per particle is approximately constant, which means that the work necessary to break a nucleus in two is very small. From medium weight onwards, the mass defect increases less rapidly than the number of particles, and this means that energy is actually gained by breaking the nucleus into two portions. The reason for this is that the comparatively weak attraction forces are now outweighed by the electrostatic repulsion between the parts of the nucleus, which increases as the size and charge of the nucleus increase.

Hence every heavy nucleus is energetically unstable against fission. It does not, in general, break up, because in order to do so it would have to pass through configurations of still higher energy; the nucleus is stable against small deformations. In order to produce the deformation leading to fission it is necessary first to elongate the nucleus, which would normally be about spherical, and then to concentrate the mass near the two ends, in order to obtain a dumb-bell shape. After this the electrostatic repulsion is probably sufficient to overcome the remaining attraction caused by the " neck " of the dumb-bell and to cause fission. The two fragments will now be thrown apart by their electrostatic repulsion.

If the charge and size of the nucleus were too big it would already be unstable for small deformations. Such a nucleus would elongate spontaneously and would then break up. Bohr and Wheeler (1939 a and b) have investigated the condition for stability under small deformations and found that the critical size of the nucleus is just a little larger than that of the heaviest known nuclei.

Hence a nucleus such as thorium or uranium is very near the stability limit. Although energy is required for a small deformation, this soon reaches a maximum and further deformation proceeds spontaneously. The energy that has to be supplied in this way represents an activation energy. If the nucleus is excited to a sufficiently high level by the capture of a neutron or by other means, it may possess sufficient energy to reach the critical state of deformation.

We know from elementary quantum mechanics that a potential barrier (i.e. the need to pass through configurations of higher potential energy) is not sufficient to prevent a reaction from taking place. By the Gamow phenomenon of penetration through a potential barrier there is always a finite chance of the transition taking place even though the energy of the system is not sufficient to reach the top of the barrier. However, the probability of penetration depends very much on the mass which is involved, and since the effective mass for the degrees of freedom which are involved in the elongation of the nucleus is of the order of the mass of the whole nucleus it is understandable that the probability is extremely small if the energy is below a critical value. Hence no fission can in practice be observed in most heavy nuclei, and even in the most unstable nuclei, uranium and thorium, the fission from their ground state is practically negligible.

As the energy of the nucleus is increased, the probability of fission rises steeply in the neighbourhood of the critical energy. The rise is not discontinuous, as it would be in classical mechanics (where the height of the potential barrier represents an absolute minimum for the required energy), but nevertheless it is very steep. The actual value of this probability is not easy to determine, because one must for this purpose investigate the distribution of the excitation energy over the various degrees of freedom and determine the probability that enough energy should happen to be concentrated in the deformation which is most suitable as a start for the fission.

This new process will now compete with radiation and neutron emission in the decay of the excited state of the compound nucleus. It was observed that, in uranium, fission can be caused by thermal neutrons, and also by neutrons of energies above about 2 Mv., but not by intermediate energies, in particular by the neutrons corresponding to the first resonance level of the uranium. In trying to explain this by the variation with energy of the partial width for the three relevant processes, Bohr and Wheeler were led to the conclusion that the most abundant isotope of uranium, ^{238}U, is comparatively stable, so that the compound nucleus formed by its combination with a slow neutron is still below the critical energy for fission. Hence the probability of fission is here negligible in comparison with radiation, and all encounters of slow neutrons with this nucleus lead, as in

the case of other nuclei, to radiative capture. As the energy of the neutron, and hence the excitation of the compound nucleus, increases, the fission probability rises steeply, until, at a few Mv., the partial width from fission becomes comparable with the total width, and hence fission takes place in an appreciable fraction of all collisions.

The effect caused by thermal neutrons is, according to Bohr and Wheeler, due to the presence in small quantities of the isotope ^{235}U. This latter is, as an isotope of odd mass, less stable than the other and gains more energy by acquiring an extra neutron. Hence it is plausible to suppose that for slow neutrons it may already be in a state above the critical energy for fission. The small abundance of this isotope is compensated by the large cross-section which can occur for thermal neutrons, because of the $1/v$ law (*cf.* § 5). Hence, at thermal energies, the rare isotope gives an appreciable contribution to the fission cross-section of the whole substance. This interpretation has recently been confirmed by direct experiments with the separated isotopes of uranium (Nier *et al.*, 1940).

REFERENCES

Bacher, 1940. *Phys. Rev.* **57**, 352.
Bethe, 1936. *Phys. Rev.* **50**, 332.
Bethe, 1937. *Rev. Mod. Phys.* **9**, 69.
Bethe, 1938. *Phys. Rev.* **53**, 39.
Bethe, 1940. *Phys. Rev.* **57**, 1125.
Bethe and Konopinski, 1938. *Phys. Rev.* **54**, 130.
Bethe and Placzek, 1937. *Phys. Rev.* **51**, 540.
Bohr, 1936. *Nature, Lond.*, **137**, 344.
Bohr and Kalckar, 1937. *Proc. Acad. Copenhagen*, **14**, 10.
Bohr, Peierls and Placzek, 1939. *Nature, Lond.*, **144**, 200.
Bohr, Peierls and Placzek, 1941. *Proc. Acad. Copenhagen.*
Bohr and Wheeler, 1939. *Phys. Rev.* **55**, 418 ; 1939 b, **56**, 426.
Bothe and Gentner, 1937. *Naturwissenschaften*, **25**, 30 and 126.
Breit, 1940. *Phys. Rev.* **58**, 506.
Breit and Wigner, 1936. *Phys. Rev.* **49**, 519.
Frenkel, 1936. *Sov. Phys.* **9**, 533.
Frisch and Placzek, 1936. *Nature, Lond.*, **137**, 357.
Goldhaber, Hill and Szilard, 1939. *Phys. Rev.* **55**, 47.
Griffiths and Szilard, 1937. *Nature, Lond.*, **139**, 323.
Hornbostel, Goldsmith and Manley, 1940. *Phys. Rev.* **58**, 18.
Kalckar, Oppenheimer and Serber, 1937. *Phys. Rev.* **51**, 450.
Kapur and Peierls, 1938. *Proc. Roy. Soc.* A, **166**, 277.
Landau, 1937. *Sov. Phys.* **11**, 556.
Lifshitz, 1938. *Sov. Phys.* **13**, 224.
Manley, Goldsmith and Schwinger, 1939. *Phys. Rev.* **55**, 39.
Nier, Booth, Dunning and Grosse, 1940. *Phys. Rev.* **57**, 546.
Oppenheimer and Phillips, 1935. *Phys. Rev.* **48**, 500.
Peierls, 1936. *Rep. Progr. Phys.* **2**, 27.
Peierls, 1940. *Ann. Rep. Chem. Soc.* **36**, 24 ; *Nature, Lond.*, **145**, 687.
Siegert, 1939. *Phys. Rev.* **56**, 750.
Walke, 1940. *Rep. Progr. Phys.* **6**, 16.
Weisskopf, 1937. *Phys. Rev.* **52**, 295.
Weisskopf and Ewing, 1940. *Phys. Rev.* **57**, 427.

ELECTRON MICROSCOPES OF HIGH MAGNIFICATION

By O. KLEMPERER, Ph.D.

§1. THE DIFFERENT TYPES OF ELECTRON MICROSCOPE OF HIGH RESOLVING POWER

Since the invention of the microscope, about 300 years ago, there has been a continuous striving to improve the resolving power in order to observe smaller and smaller objects. Glass materials and the shapes of the lenses were gradually improved in an empirical way, and the workable magnifications were gradually increased. About seventy years ago, Abbe was the first to recognize a fundamental limitation for the resolving power. According to him, the smallest distance d of two parts of an object which can be resolved with light of wave-length λ is given by

$$d = \lambda/(N \sin \alpha), \qquad \ldots\ldots(1, 1)$$

where $(N \sin \alpha)$ is the numerical aperture of the objective, N being the refractive index of the object space and α the semi-vertical angle of the cone of rays entering the objective. If the wave-length λ is decreased, this resolving distance d can be reduced. Thus if ultra-violet light is being used instead of visible light, proportional improvement in resolving power can be obtained.

Quite new prospects have been opened up by recent progress in electron physics. The fundamental theoretical investigations of L. de Broglie in 1925 and the experimental researches of G. P. Thomson in 1928 established the existence of a wave-length of the electron which was found to be extremely short in comparison with the wave-length of light. The electron wave-length measured in cm. is

$$\lambda = \frac{1{\cdot}22 \,.\, 10^{-7}}{\sqrt{V}} \times \frac{1}{\sqrt{(1 + 10^{-6} V)}}. \qquad \ldots\ldots(1, 2)$$

λ depends only upon the electron energy V, which, in equation (2), is measured in ev. Electron rays of 20 to 100 kv. energy, for instance, can easily be produced by gas discharges or by an acceleration of the emission from a hot cathode. The wave-length of these rays is about 100,000 times smaller than the wave-length of visible light. In 1926, H. Busch showed that the action of short, axially symmetrical, magnetic fields on electron rays is similar to the action of a glass lens on light rays. This discovery opened up the new science of geometrical electron-optics. A few years later, the new possibility of electron microscopy

was recognized. In 1932, M. Knoll and E. Ruska published their first experimental results on a two-stage projection-microscope focusing the electron rays by means of iron-shielded electromagnetic coils of short focal length. With these first instruments only poor pictures, of a relatively small magnification, were obtained, but the development of the new technique by E. Ruska, H. O. Müller, H. Krause and others, all working in the same technological institute at Berlin, produced gradual improvement; in 1935, the resolving power of the best optical microscope was reached; in 1937 it was surpassed by an order of magnitude. At this stage the development of the electron microscope was taken up in industry. In 1938, B. V. Borries and E. Ruska gave a first account of a technical electron microscope which they designed for the Siemens Company in

Figure 1. Commercial magnetic electron microscope.

Berlin. In the meantime, work on the magnetic electron microscope of high magnification was begun in many other laboratories, and we can only mention here the instruments developed by L. Marton in Brussels in 1934, L. C. Martin in London (1937) and by J. Hillier and A. Prebus in Toronto in 1939.

In figure 1 there is shown the detailed construction of a recent technical instrument developed at the Siemens works (Borries and Ruska, 1939a and 1939b). The arrangement of the electron lenses and the paths of the rays can be recognized from the scheme in figure 2. The electron rays are produced and accelerated to about 60 kv. in a discharge tube; they are concentrated on the object

by a condenser coil. The object is brought into the proper position by a micrometer adjustment. As in the "light" microscope, the objective forms an intermediate electron-optical image which is projected on to a fluorescent screen or to a photographic plate by means of a projection coil. In the stand of the microscope is arranged a high-speed diffusion pump maintaining high vacuum inside the microscope. Object and photographic plates are introduced from outside by means of air-locks, so that only a few minutes after their introduction the vacuum is again good enough to obtain an exposure. The electron beam only strikes the relatively tender object during the short time (order of one second) of exposure; the rest of the time the beam is deflected by a magnetic field set up in the deflection

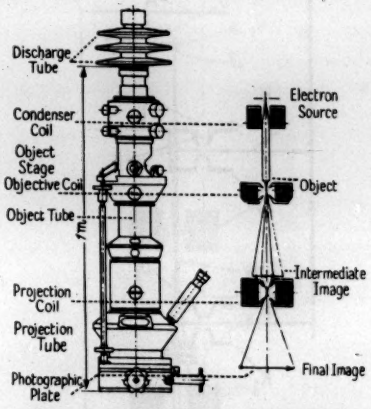

Figure 2. Scheme of the magnetic microscope.

chamber. The considerable length of the instrument, which is indicated by the scale at the left of figure 2, is necessary to produce the relatively high magnifications of about 100 times and 200 times respectively at each of the two stages of the microscope, the overall magnification being in this case 20,000 times. The magnetic lenses, which will be discussed later on in detail, have focal lengths of a few millimetres only.

Quite recently, E. Brüche, H. Mahl and others of the A.E.G. Company laboratories at Berlin have succeeded in designing the first electrostatic microscope of high magnification (Brüche, 1939; Mahl, 1939a). A schematic diagram of this instrument is shown in figure 3. Instead of the magnetic fields produced by coils, electrostatic fields produced by electrically charged diaphragms are employed as electron lenses. Electrons are emitted from a hairpin filament C which is surrounded by a cathode screen G, both electrodes being at high negative

voltage of, say, 30 kv. The electrons are accelerated towards the earthed anode A. A beam passing the stop D_1 strikes the object at B; the objective lens, consisting of the field between the diaphragms H_1, M_1 and K_1, forms an intermediate image near the stop D_2, this image being magnified by the projection lens H_2, M_2 and K_2, and focused upon the fluorescent screen S. This screen can be swung away like a flap, so that the electrons can strike a photographic plate arranged underneath the screen. The mode of operation of the electrostatic lenses is of particular

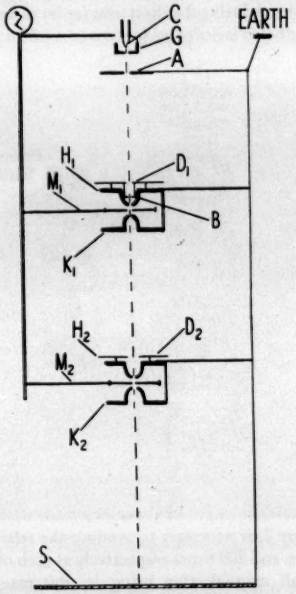

Figure 3. Electrostatic electron microscope.

interest: the external diaphragms, H_1, K_1 and H_2, K_2 respectively, are earthed, the internal diaphragms, M_1 and M_2 respectively, are at cathode potential.

The focal length of such an *einzel-lens* cannot be adjusted, as will be pointed out in the next section. In order to bring the image into focus, the object distance is varied mechanically. Whilst, however, during the time of an exposure the high-voltage supply of the magnetic microscope has to be stabilized carefully and the currents through the magnetic coils have to be kept strictly constant, no

precautions of this kind are needed for the electrostatic instrument. Moreover, the electrostatic lenses are simpler in construction and cheaper than the magnetic ones. The electrostatic microscope, however, has not quite reached the extremely high magnifications and resolving power obtained with the magnetic instrument.

The magnetic and the electric types of the two-stage electron microscope are both designed in strict analogy to the traditional glass microscope. Recently, however, two interesting attempts were made to obtain electron microscopes of high magnification on rather different and novel lines.

In figure 4 there is shown the scheme of a scanning microscope such as was developed a few years ago by M. v. Ardenne (1938a and 1938b). Electrons

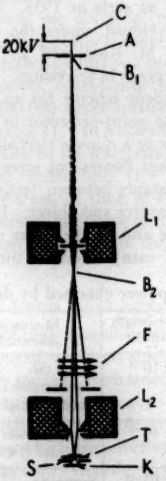

Figure 4. Scanning electron microscope.

emitted from a cathode C are accelerated through an anode diaphragm to an energy of about 20 kv. The electron beam forms the fine cross-over B_1 of about 1/10 mm. diameter when reaching its smallest cross-section. The "demagnifying" lens L_1 forms a very small image B_2 of the cross-over B_1, the cross-section of B_2 being about 100 times smaller than B_1. The image B_2 is further "demagnified" by a second lens L_2 and projected as a very fine "electron probe" of about 10^{-6} cm. cross-section on to the object T. A magnetic scanning field F, such as is generally in use for television tubes, can be employed to move the electron probe, line after line, over the object surface. The electrons transmitted through the object (or, in other constructions, those reflected from the object) are collected

by the plate K, which is connected with the grid of the input valve of an amplifier. The amplified currents are then applied to the modulator of a television tube which is scanned in synchronism with the electron probe so that the magnified picture of the microscopical object can be seen on the large fluorescent screen of this tube. From pictures obtained in this way, the resolving power of the scanning microscope appears to surpass that of the optical microscope, but the scanning microscopes are still far inferior to the highly developed projection electron-microscopes. The way was prepared for the development of the scanning microscope by methods very similar to those used in television technique. Moreover, cathode-ray tubes to detect the macro-structure of a scanned metal plate had been described by M. Knoll as early as 1935. These macroscopic scanning tubes have been further developed, under the name of "monoscope", by C. E. Burnitt (1938) and by M. Knoll and R. Theile (1939), and they are frequently used for testing purposes in television.

Another new and independent scheme for an electron microscope of high magnification is realized in the point-projector or shadow-microscope. There the object is brought very near to a minute electron source, casting an electron-shadow on to a relatively distant fluorescent screen. The magnification of the shadow image is equal to the distance between "point source" and screen divided by the distance between point source and object. If these distances are arranged to be of the order of a metre and of the order of a fraction of a millimetre respectively, magnifications of many thousand times are obtained. The main

Table 1. Resolving power obtained by different microscopes

Rays	Wave-length (cm.)	Microscope: described by	Least resolved distance (cm.)
White light (mean)	$5{\cdot}5\times10^{-5}$	Good oil-immersion: commercial	2×10^{-5}
Ultra-violet light	$2{\cdot}1\times10^{-5}$	Quartz lenses: Köhler (1938)	8×10^{-6}
20-kv. electrons	$8{\cdot}6\times10^{-10}$	Scanning microscope: v. Ardenne (1938)	$>10^{-5}$
,,	,,	Shadow microscope: Boersch (1939 b)	4×10^{-6}
,,	,,	Electrostatic two-stage: Mahl (1939 b)	2×10^{-6}
60-kv. electrons	$4{\cdot}8\times10^{-10}$	Magnetic two-stage: Borries and Ruska (1939 c); v. Ardenne (1940 a)	3 to 5×10^{-7}

problem in this scheme is the production of the point source. In their point-projector microscope, G. A. Morton and E. C. Romberg (1939) use etched tungsten points of less than 5×10^{-5} cm. radius as cathodes. The electron emission from the cathode obtained by application of sufficiently high fields is, however, not steady enough to permit photography at the very highest magnifications. H. Boersch (1939a and 1939b), in his shadow microscope, applies two stages of electrostatic lenses in order to "demagnify" the electron-optical image of a hair-pin cathode. In this way an electron source of minute dimensions is obtained, as in the method shown in the scanning microscope of figure 4.

The best resolving powers measured on images formed by the various types of electron microscope are shown in table 1.

The highest resolving power yet obtained (magnetic microscope, 1940) is thus about 3×10^{-7}. This resolving power is about 50 times better than that of the best "light" microscopes. Although this is a good achievement, the improvement is far below our expectations, considering the wave-length of the electrons used. This wave-length is 100,000 times smaller than the wave-length of the visible light (see column 1 of table 1). In Abbe's formula, equation (1), however, the resolving power depends also on the numerical aperture. The latter is of the order of one unit for optical microscopes (1·4 for oil-immersion objectives), but the largest apertures which can be used in electron microscopes are of the order of 1/1000 only. This serious limitation of the aperture is due to a spherical aberration which is surprisingly large even in the best electron-optical objective at the present time.

§ 2. THE LENSES USED IN ELECTRON MICROSCOPY

In electron microscopes of high magnification, electron lenses of very short focal length have to be employed, and iron-shielded magnetic lenses have been used since electron microscopy started. Such lenses consist, as shown, for example, in figure 5, of a coil C with many (say 2000) ampere-turns, the coil being surrounded by a shield F of mild iron ending in two pole-pieces P_1, P_2 which leave a small gap S. The magnetic field by which the electron rays are focused extends from the gap towards the axis of the coil. In order to obtain a high magnification, the object distance has to be very small. As, however, the magnetic field extends a noticeable distance along the axis, the object has to be brought into this field. Consequently, a part of the field does not contribute to the lens action, as it falls outside the space which lies between object and image.

An arrangement introduced by H. Ruska (1934b) for investigation of the characteristics of magnetic objective lenses is shown in figure 6. P_1 and P_2 are the pole pieces of the electron lens, while a fine wire gauze W lying on an aluminium foil Al serves as an object, and is arranged in a metal tube G which can be raised and lowered by means of a pulley Q in order to change the object position. The wire gauze object W is "illuminated" by the electron beam E, the rays being scattered by the aluminium foil, which is, however, so thin ($< 10^{-4}$ cm.

that it does not introduce noticeable losses of electron energy. The electron-optical image of W is projected on to the fluorescent target T. At the right hand

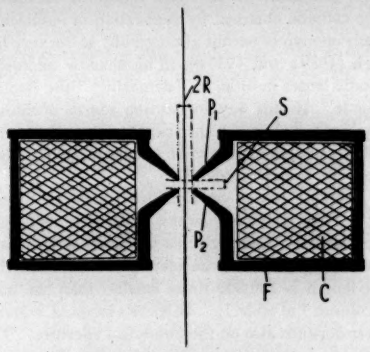

Figure 5. Magnetic electron lens.

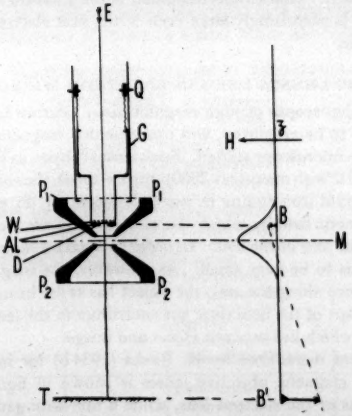

Figure 6. Investigation of an objective at different object distances.

side of figure 6 there is shown a graphical representation of the distribution of magnetic field-strength H along the axis z. Object and image are indicated by B and B'.

A measurement of the position of the foci and principal planes by means of the standard method of observing the magnifications for different positions of the object is not feasible here. A change in object position would change the fraction of the field focusing the beam, at the expense of the other fraction lying idle at the other side of the object. To obtain magnifications as high as possible, the current through the lens coil could be raised in order to increase the intensity of that fraction of the magnetic field which contributes to the focusing of the rays. The focal length of the lens is thus decreased and the object can be moved nearer to the mid-plane of the lens. This possibility of increasing the magnification by increasing the coil current is, however, rather limited. Owing to the finite value of the permeability of the pole pieces, the field intensity soon reaches a maximum, and a further increase in current only results in spreading out the field, thus

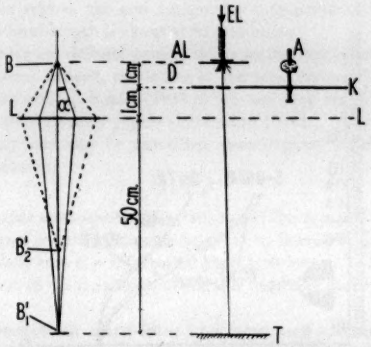

Figure 7. Arrangement for measuring spherical aberration.

reducing the focusing fraction of it. According to H. Ruska (1934 b), for any particular shape and material of pole pieces there is an optimum coil current producing the minimum focal length for a given electron energy. The maximum magnification obtainable for a given image distance is therefore limited too. According to recent calculations by E. Saguta (1939), the limiting magnification for an image produced by 100 kv. electrons at 50 cm. distance from a lens made from the best magnetic materials of today is bound to be less than 500. Thus at least two stages (objective and projection lens) will be required for all projection microscopes of high magnification.

It has already been mentioned that the resolving power of the electron microscope is limited by the spherical aberration of its objective lens. The first measurements on the spherical aberration of objective lenses were made by E. Ruska (1934 b), his experimental arrangement being explained by the scheme shown in figure 7. A beam of electrons EL strikes a diaphragm D containing

a fine aperture which is covered by the thin aluminium foil AL. The electrons are scattered by the foil in order to produce a divergent cone of rays which is intercepted by a revolving diaphragm K. This diaphragm contains a series of ring apertures which can be brought in turn into coaxial position with the cone of rays when K is rotated round the axle A. In this way beams of different apertures can be formed as shown in the scheme on the left of figure 7. There the point source B corresponds to the aperture in D, and B_1' represents the image of B which is projected on to the fluorescent target T'. The line L shows the position of the electron lens. The rays intersecting at B_1' are drawn as solid lines. If by revolving the diaphragm K another ring aperture (and thus, for instance, a wider aperture of rays) is chosen, the rays represented by the broken lines may intersect at B_2'. The image at B_2' could be detected by shifting the fluorescent target into the proper position.

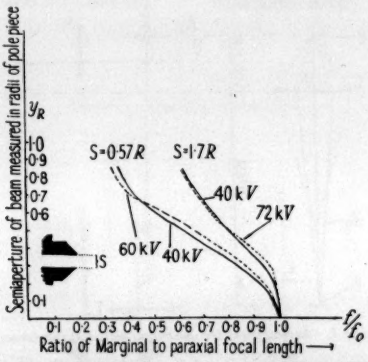

Figure 8. Spherical-aberration curves of a magnetic objective.

As an alternative, the current through the magnetic lens coil may be decreased until B_2' is focused on to the fluorescent target, which remains in a fixed position. There the focal lengths f are very approximately proportional to the square of the coil currents I^2. The latter method has been chosen by Ruska, and an example of his results is shown in figure 8. The abscissae are the ratios of marginal focal length f to the paraxial focal length f_0, and the ordinates are the semi-apertures y of the rays as measured by the semi-aperture R of the pole pieces. The particular shape of the pole pieces used is shown in the left part of the graph. The two pairs of curves correspond to the different gap-widths S of 2 mm. and 6 mm. The semi-aperture R (inner radius) of the pole pieces is 3·5 mm. in all cases. The solid curves are taken at 40 kv., the broken curves are taken with higher electron energy (60 and 72 kv. respectively). There appears to be little difference in

aberration for rays of different electron energy. In all cases the aberration is surprisingly large. Even with the larger gap width, producing definitely better results, the focal length is decreased by at least 30 per cent if the aperture of the rays is opened up to about half the width of the pole pieces.

In some recent systematic investigations on large iron-shielded magnetic lenses, H. Becker and A. Wallraff (1938) found a maximum spherical aberration for magnetic gap-widths of the order of 0·2 R ($2R$ = inner diameter of the iron shield). For larger gaps (in agreement with Ruska's results), and also for smaller gaps, the aberration decreases. Iron-shielded lenses with small magnetic gaps, however, cannot be used for microscopical purposes as their stray fields are comparatively weak and consequently their focal lengths are relatively large even for extremely large coil-currents. Thus it is advisable to use relatively wide gaps, and consequently gap-widths of the order of the pole-piece radius R are used in practice. In this region, the coil current (as a function of the gap width) to produce a given focal length is about at its minimum.

It appears there are definite natural limits set to improvements in the design of magnetic lenses. Indeed, the lenses of the most recent objectives described by B. v. Boerries and E. Ruska (1939 a) are not very much better than those described in H. Ruska's early papers (1934 a and 1934 b). The primary spherical aberration of any lens may be described quantitatively by the radius h of the disc of least confusion:

$$h = C\alpha^3, \qquad \ldots\ldots (2, 1)$$

α being the angular semi-aperture (in radians) of the rays at the object side, and C being a constant for the particular lens. B. v. Borries and E. Ruska (1939 c) find experimentally that $C = 150$ cm. for their best lenses. The meaning of this value with respect to the maximum obtainable resolving power will be explained later on.

Recently, electrostatic *einzel-lenses* have been used successfully by H. Mahl (1939 b) and by M. v. Ardenne (1939 c, 1940 a) as an alternative to the magnetic objectives. According to H. Boersch (1939 b) three-diaphragm *einzel-lenses* can be used as objectives if the spacings of the diaphragms are sufficiently large in comparison with the aperture radii. In this case the focal length of the lens is relatively small, and so is the spherical aberration. According to M. v. Ardenne's measurements (1939 c), the state of correction of *einzel-lenses* of this kind is of the same order as that of one of the best magnetic objectives.

In figure 9 is shown the *einzel-lens* now used in the A.E.G. laboratories by H. Boersch (1939 b), H. Mahl (1939 b) and E. Brüche and E. Hagen (1940). Between two external diaphragms, which are connected to the earthed anode of the system, there is the middle diaphragm M fixed to the insulator J and connected through the insulated lead H with the high-tension cathode potential. B is a demountable diaphragm with fluorescent screen, on the lower part of which the object carrier is fixed. The metallic case K connecting and aligning the two external diaphragms fits on to the microscope tube V. The whole system is

arranged in the evacuated tube R. The lens is made from highly polished chromium-nickel steel, and all parts are carefully rounded so that the danger of sparking is reduced to a minimum. Spurious external magnetic fields are shielded off by a double magnetic shield consisting outside of mild iron and inside of Mu-metal. The three apertures are about equal in width (radius 0·5 mm.), their spacing being about 4 aperture-radii.

If the middle diaphragm is fixed at cathode potential, the focal length ($f \approx 5$ aperture-radii) is constant, and does not depend on the potential difference between cathode and anode which defines the energy of the electrons. It has been

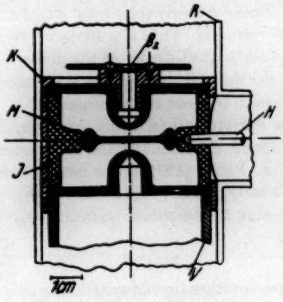

Figure 9. Electrostatic objective.

mentioned already that this immunity to voltage fluctuations is a great advantage of the *einzel-lens* over the magnetic lens, for which the accelerating anode-voltage has to be stabilized carefully. On the other hand, this rigid constancy of the focal length of the electrostatic lens requires a mechanical device for shifting the object carrier so that the object can be brought into focus.

As an alternative, the focal length of the *einzel-lens* can be made adjustable, if the middle diaphragm of the lens is kept at a variable voltage intermediate between anode and cathode potential. This is done, for instance, in M. v. Ardenne's microscope (1939c and 1940), in which the object need not be shifted for focusing. The advantage, however, of independence of voltage fluctuations is lost.

§3. FORMATION OF THE IMAGE IN THE MICROSCOPE; LIMITATIONS ON THE RESOLVING POWER

Numerical apertures used in electron microscopy are very small in comparison with those to which we are accustomed in light microscopy, since they are of the order of 1/1000 only. In light microscopy the aperture of rays is limited by suitable stops or even by the metal frame holding the first glass lens of the system. In electron microscopy, however, the apertures of rays passing the object are given

by the apertures of rays coming from the condenser lens. A scheme of rays as it occurs, for instance, in the first stages of the microscopes of figures 2 and 3 is shown in figure 10. In this figure the positions of electron source, condenser lens, object, objective lens and intermediate image are indicated. Bundles of

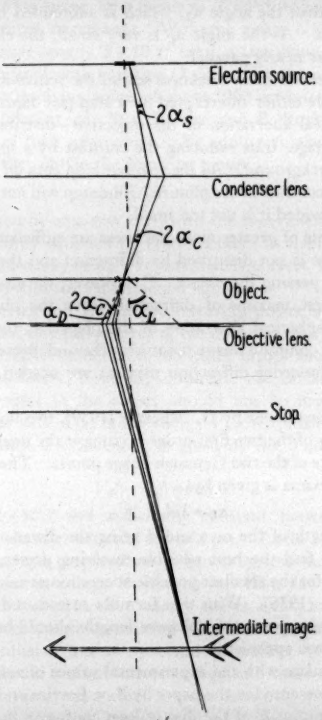

Figure 10. Formation of the image in an electron microscope.

rays leaving the source with the small semi-divergence α_S are focused by the condenser on to the object, the angle of convergence α_C being given by α_S and by the mutual distances of source and object from the condenser. The rays passing the object form a diffraction pattern which is focused by the objective, thus producing the image. The process is quite analogous to the image formation

in the optical microscope which is explained by Abbe's theory. As shown in figure 10, the numerical aperture of the electron microscope depends upon α_C, the angle of convergence of rays from the condenser. To α_C has to be added α_D, the angle of diffraction of the electron waves at the object, but $(\alpha_C+\alpha_D)$ is still very much smaller than the angle α_L, which is subtended by the free aperture of the objective lens. As the angle α_C is very small, the electrons transmitted through the object are nearly parallel.

Object parts of greater mass-thickness scatter the transmitted electrons. The scattered electrons are either intercepted by a stop (see figure 10), or they are—owing to the spherical aberration of the objective—distributed over relatively large areas of the image, thus reducing the contrast by a more or less uniform illumination of the background. As the aperture stop cuts off the stray electrons, its width defines the contrast in the picture. The stop will not affect the sharpness or the definition, provided it is not too small.

If the object points of greater mass-thickness are sufficiently far distant from each other, the image is not disturbed by diffraction and the rays are still sufficiently parallel after passing the object. If, however, the structure of the object is small, even the first maxima of diffraction leave the object at a noticeable angle. Due to the spherical aberration of the objective, these maxima are no longer focused into a common image point with the undeflected rays. According to Abbe, the two first-order diffraction maxima are needed to produce a true image.

According to an argument by O. Scherzer (1939), two object points cease to be resolved if the rays of the two first-order maxima at the image plane are further apart than the distance of the two Gaussian image points. The angle of diffraction for the first-order maxima is given by

$$\alpha_D=\lambda/d, \qquad \ldots\ldots(3,1)$$

λ being the wave-length of the rays and d being the distance of the two object points. In order to find the best possible resolving power, O. Scherzer uses a theoretical formula for the smallest possible aberration as calculated in an earlier paper by R. Rebsch (1938). With this formula introduced in equation (3, 1) it follows that a resolving power of 100 wave-lengths should be obtainable. This estimate seems to us too optimistic.

We prefer to calculate with the experimental values of actual aberration constants such as those presented in the paper by B. v. Borries and E. Ruska (1939 c). There we obtain the radius h' of the disc of least confusion in the image plane as given by

$$h'=\frac{y'}{y}\,.\,C\,.\,\alpha^3. \qquad \ldots\ldots(3,2)$$

Here C is a constant (see equation (2, 1)), α is the semi-divergence of the rays at the object side and y'/y is the magnification. The distance of the Gaussian image points is then

$$d'=(y'/y)d \qquad \ldots\ldots(3,3)$$

and the condition for the minimum distance of the object points is

$$(y'/y)d_{\min} = (y'/y)C\alpha^3 = (y'/y)C(\lambda/d_{\min})^3, \quad \ldots\ldots(3, 4)$$

i.e.

$$d_{\min} = \lambda(C/\lambda)^{1/4}. \quad \ldots\ldots(3, 5)$$

$d_{\min}$ of equation (3, 5) is larger by a factor of $(C/\lambda)^{1/4}$ than the resolving power of an optically corrected microscope. With the value $C = 150$ cm., as given in § 2, and for a wave-length 5×10^{-10} cm., a resolving power $d_{\min}$ of 750 wave-lengths is calculated with equation (3, 5), whilst the experimental results given in table 1 give a resolving power of about 1000 wave-lengths.

In a somewhat different way B. v. Borries and E. Ruska (1939 c) calculate the resolving power of their microscope. By adding an aberration error h and a diffraction error d, they obtain the resolving power

$$d_r = \sqrt{(h^2 + d^2)}. \quad \ldots\ldots(3, 6)$$

The value of h is given by equation (2, 1) with the experimental figure of C, and the value of d is given by equation (3, 1) with α taken as the condenser aperture. Although this procedure may be only approximately correct, the results appear to be in good agreement with the experimental figures.

According to Borries and Ruska, the resolving power seems to be entirely limited by the spherical-aberration error of the objective, and in comparison with its magnitude the relative influence of all other errors should be very small. As other authors emphasize, however, chromatic aberration due to straggling of the electron energies in the object should not be neglected. The mean energy V_x of the electrons of an initial energy V_0 after passing an object of mass thickness sx (s = density, x = thickness) is given by the Thomson-Whiddington formula:

$$V_0^2 - V_x^2 = b\,sx, \quad \ldots\ldots(3, 7)$$

where b is about 2×10^{-5} and sufficiently constant between $V = 20$ and 70, V being measured in electron-kilovolts and x in cm. As the mass thickness of the usual electron-microscope preparations is generally less than $sx = 10^{-5}$ g./cm^2, the mean loss of electron energy should be less than one per mil for 60-kv. electrons. The small loss of electron energy causes a very small change in focal length of the electron lens, and the resulting disc of least confusion should be much smaller than the confusion due to spherical aberration. On the other hand, equation (3, 7) applies for a mean loss of energy, and there are still other losses which are difficult to estimate and which in some cases may spread over a relatively wide band. The larger energy losses are in all cases comparatively rare, but under these conditions it is difficult to give an accurate estimate of the magnitude of the chromatic error (v. Ardenne, 1939 c). Moreover, there are particular effects produced by the chromatic aberration which modify the intensity distribution in the image. According to a discussion of the problem by J. Hillier (1939), the intensity should show a maximum just outside and a minimum just inside the image of any edge of the object. Under present circumstances, the chromatic error (radius of disc of least confusion) might be guessed to be of the order of

one-tenth as serious as the spherical aberration error, but the situation, however, will change greatly when some correction of the spherical aberration becomes practicable.

Spherical aberration sets an all-important limitation not only for existing projection microscopes, but also for scanning microscopes and for shadow microscopes. In the scanning microscope the smallest available cross-section of the electron probe is obviously limited by aberration and diffraction (Scherzer, 1939). In the shadow microscope, according to Boersch (1939 b), the aberration produces no blurring but only a barrel distortion of a perfectly sharp picture. We must refrain from going into detail, but we mention the result of an analysis by O. Scherzer (1939) according to which a characteristic combination between diffraction and aberration limits the resolving power in the shadow microscope to practically the same value as in the other microscopes.

With a given resolving power the magnification of a microscope has to be kept at such a value that all the details of the object structure can be recognized. Excessive magnifications are of no use, as they do not reveal new details, but increase the structureless "empty" areas only. On the other hand, it may be of a definite advantage to keep the electron-optical magnification as small as possible and to arrange a light-microscopical post-magnification. For a given current density at the object, the time of exposure of the photographic plate can be reduced as soon as smaller magnifications, and thus larger current densities, are obtained. If fluorescent screens are used, the picture could be observed by eye directly or through a light-microscope. The aperture of the light-microscope, however, is generally large in comparison with the aperture of the eye ($\alpha \simeq 0{\cdot}01$). An increase in brightness of the final image results if the post-magnification is chosen to be smaller than the ratio of the two apertures. Practically such a gain is easily obtained, as the apertures of light-microscopes are generally 10 to 100 times larger than the aperture of the eye.

According to v. Ardenne (1939 d), the magnification to obtain the best current densities on the screen or plate is only slightly larger than the minimum value given by

$$y'/y_{\min} = h'_{\min}/d, \qquad \ldots\ldots(3, 8)$$

d being the resolving power of the microscope and $h'_{\min}$ being the smallest distance that is still resolved by the photographic plate or by the fluorescent screen.

According to M. v. Ardenne (1939 f), the resolving power of suitable commercial photographic plates is about 2×10^{-3} cm., whilst 3×10^{-4} cm. have been resolved with gelatine-free Schumann plates. This gain by a factor of 6 in resolving power may reduce either the time of exposure or the current density at the object by a factor of 36.

The resolving power of ordinary fluorescent screens with particle sizes of 10^{-3} to 10^{-4} cm. is not better than 10^{-2} cm. With preparations of artificially reduced particle sizes, M. v. Ardenne (1939 a) measured resolving powers of 10^{-3} cm. The luminous efficiency of fluorescent substances of very fine grain,

however, is relatively small. An early patent-proposal of R. Rüdenberg (1932) suggests the use of large single crystals. With such single crystals of about 1 mm^2 surface and a thickness ground down to 3×10^{-3} cm., v. Ardenne (1939a) obtained relatively high luminous efficiency with resolving powers better than 10^{-3} cm. The resolving power of fluorescent screens is quite generally limited by the scattering of the electrons in the fluorescent substance. In addition, there exists another limitation which follows from simple geometrical considerations on the depth of focus. According to this optical limit, the smallest resolved distance is

$$h'_{\min} = \frac{x \sin \alpha}{N}, \qquad \ldots\ldots(3, 9)$$

x being the thickness and N the refractive index of the screen ($N \simeq 2{\cdot}3$ for ZnS) and $\sin \alpha$ the numerical aperture of the observing light-microscope.

§ 4. THE USE OF ELECTRON MICROSCOPES OF HIGH MAGNIFICATION

Magnetic as well as electrostatic microscopes have reached a sufficiently advanced stage of development for their practical application in material testing and in biology to be well considered. In figures 11 to 15 we show some examples

Figure 11. Synthetic kaolin.

of the efficiency of these new instruments: figures 11 to 13 were taken with a magnetic microscope, figures 14 and 15 with an electrostatic microscope. D. Beischer and F. Krause (1938), M. v. Ardenne and D. Beischer (1940) and

Figure 12. Blood corpuscle.

Figure 13. Tobacco mosaic virus.

H. Friess and H. O. Müller (1939) described the investigation of the microstructure of inorganic materials, especially of precipitates and of small particles in dusts and smokes. The investigation of fine colloidal particles was attacked by F. Krause (1938). Some papers on the structure of cement clays and ceramics were published by H. O. Müller (1938) and by M. Eitel, H. O. Müller and

Figure 14. Crystallites of chemically precipitated magnesium oxide.

Figure 15. Bacillus tertius amylobacter with flagellates. Age 4 days.

O. E. Radczewski (1939). In metallurgy the investigation of thin oxide films providing protection from corrosion has been studied by H. Mahl (1940).

The most important applications of the new microscopes are undoubtedly concerned with medicine and microbiology. The beautiful photographs of bacteria reveal new details which could not be seen by the optical microscope.

Of the great number of papers showing pictures of bacteria we quote those by B. v. Borries and E. Ruska (1938 b), B. v. Borries, E. Ruska and H. Ruska (1938), H. Mahl (1938 b) and by G. Piekarski and H. Ruska (1939 a and b). Moreover, the virus research will probably profit greatly from electron microscopy, as the size of a virus is beyond the resolving power of light-microscopes, but it is clearly observed in electron-microscopical pictures by C. A. Kausche, E. Pfankuch and H. Ruska (1939). The investigations by C. Wolpers and H. Ruska (1939) on blood corpuscles offer another example of the use of the new microscope in medicine.

The technique for preparing electron-microscopical objects has been described in detail by H. Ruska (1939). More robust objects like smoke particles, or precipitates, etc., can be brought into the microscope supported on a very thin molybdenum wire gauze. Organic textures or bacteria, etc., are usually mounted on very thin films of collodion or nitrocellulose. If these films are thin enough, the amount of scattering and spread of velocity caused by them does not interfere with the picture. To produce very thin films (10^{-6} cm.), a small drop ($\sim$0·01 g.) of a 1·5 per cent solution of collodion in amyl acetate is dropped on amyl-acetate-saturated water. The film spreading over the surface is taken up and dried on a metal diaphragm. Films of this kind are thinner than the length of a collodion molecule. Objects as small as 10^{-19} gram have been detected on these films, showing that the scattering power of even these minute particles was considerable in comparison with the scattering by the film. The film will stand an electron bombardment of 0·1 to 1 ma./mm^2, this current density being sufficient for a bright picture of a magnification of 20,000.

The problem of keeping the current density as low as possible is of great importance for the investigation of tender organic objects. For a photographic picture the current density can be kept low when long times of exposure are used. For adjustment purposes, however, the fluorescent screen is necessary, and the least brightness required to observe whether the image is in focus is about 0·1 ft.-candles. This is obtainable under the best conditions by about 10^{-10} amp./cm^2. Zinc sulphides, willemites or zinc beryllium silicates are suitable screen materials. Good screens of fine grain (about 1 μ) show a luminous efficiency of the order of 5 candles per watt when bombarded by 5-kv. electrons. This efficiency rapidly decreases with increasing electron energy if particular precautions are not applied. The surface of the fluorescent material is charged by the bombardment of the primary electrons, and it is discharged by conduction to the target and by secondary electron emission. The light output depends largely upon the equilibrium potential reached, as this defines the energy of the electron striking the screen (see, for instance, S. T. Martin and L. B. Headrick (1939)). At low-current densities and with metal targets this equilibrium potential can be kept sufficiently low. K. Scherrer and R. Rubsaat (1937) observed efficiencies of more than 10 candles per watt with 60-kv. electrons and minimal currents on ZnS screens. If this figure is combined with the above value for the minimum

required brightness, a value of about 10^{-10} amp./cm^2 is calculated for the minimum required current-density at the electron image.

From the minimum current-density at the image the necessary current-density at the object can be derived. The ratio of the two densities is inversely proportional to the square of the magnification. With the minimum magnification defined by equation (3, 8), the necessary current-density at the object is obtained as

$$j_{min} = j'_{min}(h'_{min}/d)^2 \text{ amp./cm}^2 \qquad \ldots\ldots(4, 1)$$

With the above value $j'_{min} = 10^{-10}$ amp./cm^2, and with a resolving power of the microscope of $d = 5 \times 10^{-7}$ cm., assuming the smallest distance resolved by the fluorescent screen to be $h'_{min} = 10^{-3}$ cm., a minimum current-density through the object of $j_{min} = 4 \times 10^{-4}$ amp./cm^2 is required. M. v. Ardenne (1939 b) compares this value j_{min} with a critical electron-density of 3×10^{-6} coulomb/cm^2, being just sufficient to kill living substances by its ionization effects. According to this conclusion it should not be possible to observe living substances under an electron microscope of high magnification.

It remains still to indicate some possibilities of future development by which the use of the electron microscope might be enlarged greatly. Due to the very small aperture of the electron rays, the electron microscope shows a surprisingly large depth of focus. In an early patent proposal, E. Ruska suggested the possibility of inspecting relatively large depths of the object and of developing an electron stereo-microscopy. Hopeful experiments in this direction have recently been made by M. v. Ardenne (1940), who introduces in his microscope a particular object carrier which can be tilted by a few degrees between two successive exposures. An inspection of the two corresponding photographs under a stereoscope produces a vivid impression of solidity.

Another possible development lies in the application of dark-ground illumination methods to electron microscopy. There the objective aperture is relatively large, the direct rays from the object are cut off by a central stop and only scattered electrons pass the objective. B. v. Borries and E. Ruska (1939 b) obtained some photographs with this method which show, however, some lack of sharpness and definition in comparison with the pictures obtained in ordinary electron microscopy. This would be expected, as the larger objective apertures are liable to larger disturbances due to spherical aberration. Moreover, the picture is formed by deflected electrons, but the loss of energy of deflected electrons is always larger than that of the undeflected ones. Thus the chromatic error too is relatively large. On the other hand, M. v. Ardenne (1940) applied dark-ground illumination successfully and obtained resolving powers down to 5×10^{-7} cm. There the objects extended freely into the aperture of an object carrier. According to v. Ardenne, objects when mounted on foils are unsuitable for dark-ground illumination methods, as even the thinnest foils produce illumination by scatt ring.

Dark-ground illumination as used in the well-known light ultra-microscope of Siedentopf and Zsigmondy should be of particular interest. In electron

ultra-microscopy it should be possible to observe particles which are smaller than the resolving power of the instrument. M. v. Ardenne (1939 d) discusses in this connection the possibility of viewing single atoms and the study of their distribution in the object plane. There are, however, still great practical difficulties, to mention only the time of exposure, which would have to be increased more than 1000 times if ultra-microscopical methods were to be introduced.

On the other hand, it might be possible to improve the resolving power of the ordinary electron microscope with direct illumination by an improvement of the lenses leading to an application of larger numerical apertures. O. Scherzer (1939) discusses in this connection the possibility of correcting spherical aberration by introducing space charges into the lens. A patent proposal by F. H. Nicoll in 1936 discusses the introduction of an electron mirror into the microscope. As it is possible to make mirrors with negative aberration, a useful opportunity of correcting the mirror-microscope is given. The obvious method of improving the resolving power by using appreciably greater electron energies, and thus shorter wave-lengths, is probably ruled out, as the difficulties of handling these extreme voltages would be very great. The evident difficulties involved in all these proposed methods of improving the electron microscope should, however, by no means produce the impression of final limits. On the contrary, the whole subject of electron microscopy is still in its infancy and, remembering the gradual but enormous development which was experienced in the history of the light-microscope, we may well expect that further research on the subject may place very much higher resolving powers within our reach. Apart from all future developments, however, the electron microscope, as it stands, offers a new and powerful tool which already may find a useful application in many research institutions, in material testing and in biological laboratories or in hospitals.

REFERENCES

The above Report was concluded in April 1940. The following list refers to literature published mainly during the last two years. Excepting three earlier papers containing important pioneer work, we have refrained from quoting earlier literature, as it is catalogued with sufficient completeness in the summarizing monographs by O. Klemperer (1939) and by L. M. Myers (1939).

v. Ardenne, M., 1938 a. *Z. tech. Phys.* **19**, 407.
v. Ardenne, M., 1938 b. *Z. Phys.* **108**, 338.
v. Ardenne, M., 1938 c. *Z. Phys.* **109**, 553.
v. Ardenne, M., 1938 d. *Z. Phys.* **111**, 152.
v. Ardenne, M., 1939 a. *Z. tech. Phys.* **20**, 235.
v. Ardenne, M., 1939 b. *Z. tech. Phys.* **20**, 239.
v. Ardenne, M., 1939 c. *Z. tech. Phys.* **20**, 289.
v. Ardenne, M., 1939 d. *Z. Phys.* **112**, 744.
v. Ardenne, M., 1939 e. *Z. Phys.* **113**, 257.
v. Ardenne, M., 1939 f. *Z. Phys.* **114**, 379.
v. Ardenne, M., 1940. *Z. Phys.* **115**, 339.
v. Ardenne, M. and Beischer, D., 1940. *Z. angew. Chem.* **53**, 103.

BECKER, H. and WALLRAFF, A., 1938. *Arch. Elektrotech.* **32**, 664.
BEISCHER, D. and KRAUSE, F., 1938. *Z. angew. Chem.* **59**, 331.
BOERSCH, H., 1939 a. *Naturwissenschaften*, **27**, 418.
BOERSCH, H., 1939 b. *Z. tech. Phys.* **20**, 346.
v. BORRIES, B. and RUSKA, E., 1938 a. *Wiss. Veröff. Siemens-Konz.* **17**, 99.
v. BORRIES, B. and RUSKA, E., 1938 b. *Wiss. Veröff. Siemens-Konz.* **17**, 107.
v. BORRIES, B. and RUSKA, E., 1938 c. *Z. tech. Phys.* **19**, 402.
v. BORRIES, B. and RUSKA, E., 1939 a. *Naturwissenschaften*, **27**, 281.
v. BORRIES, B. and RUSKA, E., 1939 b. *Naturwissenschaften*, **27**, 577.
v. BORRIES, B. and RUSKA, E., 1939 c. *Z. tech. Phys.* **20**, 225.
v. BORRIES, B. and RUSKA, E., 1939 d. *Z. wiss. Mikr.* **56**, 317.
v. BORRIES, B., RUSKA, E. and RUSKA, H., 1938. *Klin. Wschr.* **17**, 921.
BRÜCHE, E. and HAGEN, E., 1939. *Naturwissenschaften*, **27**, 809.
BURNITT, C. E., 1938. *R.C.A. Review*, **2**, 414.
BURTON, E. F., HILLIER, J. and PREBUS, A., 1939. *Phys. Rev.* **56**, 1171.
EITEL, W., MÜLLER, H. O. and RADCZEWSKI, O. E., 1939, *Ber. dtsch. keram. Ges.* **20**, 165.
FRIESS, H. and MÖLLER, H. O., 1939. *Die Gasmaske*, **11**, 1.
HILLIER, J., 1939. *Canad. J. Res.* **17**, 49.
KAUSCHE, G. A., PFANKUCH, E. P. and RUSKA, H. 1939. *Naturwissenschaften*, **27**, 292.
KLEMPERER, O., 1939. *Electron Optics* (Cambridge).
KNOLL, M. and RUSKA, E. 1932. *Ann. Phys., Lpz.*, **12**, 607 and 641.
KNOLL, M. and THEILE, R., 1939. *Z. Phys.* **113**, 260.
KÖHLER, A., 1938. *Naturwissenschaften*, **21**, 165.
KRAUSE, F., 1938. *Radiologica*, **3**, 122.
MAHL, H., 1939 a. *Naturwissenschaften*, **27**, 417.
MAHL, H., 1939 b. *Z. tech. Phys.* **20**, 316.
MAHL, H. 1940. *Z. tech. Phys.* **21**, 17.
MARTIN, L. C., 1938. *Nature, Lond.*, **142**, 1062.
MARTIN, L. C., 1939. *J. R. Micr. Soc.* **59**, 217.
MARTIN, L. C., PARNUM, D. H. and SPEAK, G. S., 1939. *J. R. Micr. Soc.* **59**, 203.
MARTIN, S. T. and HEADRICK, L. B., 1939. *J. Appl. Phys.* **10**, 117.
MORTON, G. A. and ROMBERG, E. G., 1939. *Phys. Rev.* **56**, 705.
MÜLLER, H. O., 1938. *Electrotech. Z.* **59**, 1189.
MYERS, L. M., 1939. *Electron Optics* (London).
NUTTALL, A. K., 1938. *Metrop. Vick. Gaz.* **17**, 256.
PIEKARSKI, G. and RUSKA, H., 1939 a. *Klin. Wschr.* **18**, 383.
PIEKARSKI, G. and RUSKA, H., 1939 b. *Arch. Mikrobiol.* **10**, 302.
PREBUS, A. and HILLIER, J., 1939. *Canad. J. Res.* **17**, 49.
REBSCH, R., 1938. *Ann. Phys., Lpz.*, **31**, 551.
RUSKA, E., 1934 a. *Z. Phys.* **87**, 580.
RUSKA, E., 1934 b. *Z. Phys.* **89**, 90.
RUSKA, H., 1939. *Naturwissenschaften*, **27**, 287.
SAGUTA, E., 1939. *Electrotech. J., Tokyo*, **3**, 271.
SCHERRER, K. and RÜBSAAT, R., 1937. *Arch. Elektrotech.* **31**, 821.
SCHERZER, O., 1939. *Z. Phys.* **114**, 427.
WOLPERS, C. and RUSKA, H., 1939. *Klin. Wschr.* **18**, 1077 and 1111.

NEW LENS SYSTEMS

By H. W. LEE, B.A., F.Inst.P.

§1. ASPHERIC SURFACES

For a number of years aspheric surfaces have been employed in condensers to correct spherical aberration over a large aperture. Such surfaces as are produced are usually only approximations to the ideal form, as they have to be ground and polished by mechanical means. More accurate surfaces are produced in the collimators and objectives of spectrographs by "figuring" in the polishing process; single lenses may be employed here, as achromatism is not necessary at the relatively small apertures employed (f/11), since the photographic plate can be inclined to the axis provided the spherical aberration can be corrected.

Astronomical reflectors are, of course, paraboloidal; the surface is first made spherical and then polished locally till the correct figure is attained, as shown by examination of the image of a star. (With the 200-inch mirror the amount of material is too great for removal by polishing alone, and some approximation to the paraboloid must be made in the grinding process.) An alternative, which is becoming more widely used for all but the largest mirrors, is to make the reflector of exact spherical form and to correct the aberration by means of a suitably shaped plate placed in front of it (see §2, Telescopes). The surface of this plate departs too far from the flat to be produced by local polishing,* so that some approximation to the required form must be made by grinding, which necessitates preliminary calculation.

Among the latest patents for aspheric surfaces are those of W. Merté (Zeiss, 1935 b) for a single (or double) deformed meniscus lens for photography; Kodak (1934) for a single lens nearly approximating to a parabolic cylinder, having an aperture of f/1 for use in sound recording; Zeiss (1936 a) and Kodak (1936) use toric surfaces in the imagery of a straight line (illuminated slit and lamp filament respectively). If a long cylinder is used to form an image of a straight line, the image will be considerably curved because the axial part will be formed by refraction in a normal section of the cylinder, which is circular, while the ends of the slit will be refracted by an oblique section which is elliptical, having greater vertex curvature than the circular section and consequently greater refracting power. To avoid this the generating lines of the cylinder must be curved concave to the object, so that its surface is toric. The theory of the Kodak lens,

* The reduction in optical path, in refraction, when a thickness of glass t is removed, is $(n-1)\,t$; the increase in path, in reflection, by removing the same thickness, is $2t$; thus for a given alteration in path in refraction approximately four times as much material must be removed as in reflection.

which is used in sound recording, is given by J. H. McLeod and F. E. Altmann (1938). H. W. Lee (1939) has pointed out and corrected an error in their theory of refraction through a cylinder.

Attempts have been made to use aspheric surfaces in more complicated optical systems. In high-aperture microscope objectives the surfaces are too small for local polishing (but see § 3, Microscopes for a reflecting objective). Aspheric surfaces are not amenable to large-scale production in photographic lenses, and they are difficult to test accurately during manufacture. The mathematician can express the form of surface he requires in terms of some system of coordinates, e.g. Cartesian or polar, but the manufacturer requires some geometric property by which he can generate the surface. The conic sections have simple geometric properties which allow of their production mechanically with some degree of accuracy, so H. W. Lee (1935 a) suggested the use of ellipsoidal surfaces in two types of photographic lens of extremely high aperture. Since the basic types of lens proposed already have large apertures with spherical surfaces only, it was found that part only of the surfaces need be ellipsoidal, the greater part remaining spherical. Djian (1936) made a similar proposal for large-aperture objectives, and A. Warmisham (1938 a) has also proposed surfaces partly spherical and partly aspheric, giving a polar equation for the latter; presumably this is a prescription favourable for manufacture.

§ 2. TELESCOPES

The chief development during recent years in telescope optics has been in the increasing use of the Schmidt reflector. B. Schmidt (1931) suggested that the aberration of a spherical mirror could be removed by placing in front of it a

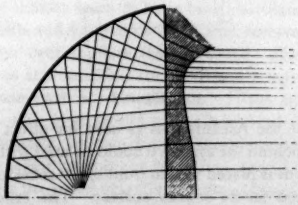

Figure 1.

refracting element having one surface suitably figured. Figure 1 (due to H. R. Schalz, 1932) shows the form of the correcting surface for a hemispherical reflector. The advantage of the Schmidt arrangement is the greatly enlarged field as well as aperture. The parabolic reflector suffers from coma, which spoils the definition of images off the axis and limits the field and aperture that can be usefully employed, since coma defect varies as the square of the aperture and directly as the field. With large reflectors the field is limited to a fraction of a degree and the aperture

to a maximum of f/3·5. F. E. Ross (1935) has designed correcting lenses for reflectors; an uncemented doublet, of zero power, placed close to the focal plane, corrects the coma and extends the field to a few degrees, provided the photographic plate is flexed to the curvature of the field. Such a corrector is to be provided for the 200-inch mirror now being polished. With the Schmidt corrector, apertures up to f/1·5 or more with fields of 20 degrees are possible. The correcting plate is placed at the centre of curvature of the mirror because the coma of a spherical reflector is zero for a stop at its centre. The Schmidt telescope cannot be used with an eyepiece, but must have the photographic plate at its focus and suitably bent to the curved field, which has a radius of curvature equal to the focal length of the mirror. Two Schmidt telescopes are being provided as finders for the 200-inch reflector, one having a focal length of 18 inches and an aperture of f/2, the other a 48-inch f/2·5. A Schmidt telescope is also being provided for the McDonald telescope; the mirror has an aperture of 122 mm. and focal length of 180 mm.; the corrector an aperture of 92 mm. (f/2).

H. W. and L. A. Cox (1939) and R. C. Waland (1938) have described methods of making and testing the plates.

Other reflecting telescopes recently patented are those of Wild (1937) and Askania Werke (1934); details are lacking in the specifications, but the writer has

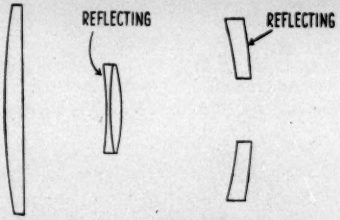

Figure 2.

been able to examine the Askania lens as used for direct photography. As the focus is outside and behind the system it could be used with an eyepiece (figure 2). A simple collective lens is placed well in front of the reflector, which is a dispersive meniscus lens silvered on its convex surface. The reflector is over-corrected for spherical aberration and rather more than compensates for the under-correction of the refractor. Between these two is a cemented doublet meniscus lens, silvered on its concave surface, which is dispersive in effect. The twice-reflected light passes out through a hole in the large reflector. Optically the system is telephotographic, with a back focus of about one-fifth the focal length, and is well corrected for astigmatism as well as spherical and chromatic aberrations and coma, and well covers a field of 10 degrees at an effective aperture of f/3·5.

Zeiss (1936 b) and H. W. Lee (1937) have sought to extend the aperture of refracting objectives by using two separated components, at least one of which is

necessarily compound; in the former the components are respectively collective and dispersive, in the latter both collective.

It is well known that the degree of achromatism to be obtained with two glasses is limited by the fact that the ratio of the dispersions of the glasses varies with the spectral region chosen, so that if a lens is made to have a common focus for rays from two points in the spectrum, the focus for a ray from a third point will be different; the residual colour is known as secondary spectrum. Some improvement can be obtained by using three glasses suitable chosen, so as to bring rays from three points of the spectrum to coincidence; also special glasses (flints) have been made with the ratio of dispersion to those of crown glass nearly constant throughout the spectrum. But the refractive power and dispersion are both so little different from those of crown that it is impossible to make lenses of large aperture from them on account of the steepness of the surfaces involved. Improvement has been sought in the use of minerals, notably quartz and fluorite, and a quartz-fluorite-water combination gave still better results (Bricout, 1931). Unfortunately fluorite is obtainable only in small pieces. Recently lithium fluoride and lithium bromide have been proposed as optical media, and D. C. Stockbarger (1937) has developed a technique for the production of large crystals. It is found that quartz and lithium fluoride have very favourable partial dispersions, with a useful difference in refractive index and dispersion. C. H. Cartwright (1939) has computed a doublet and a triplet using lithium fluoride as the crown and quartz as the flint component, achromatized from 2300 to 14,000 A., with an error in focus for any colour never greater than one-quarter of one per cent. The triplet employs two quartz elements, respectively right- and left-handed, so that rotation of the plane of polarization is practically eliminated; the quartz lenses enclose and protect the lithium fluoride element. The refractive indices seem to have been calculated by extrapolation from the Hartmann formula and so must be taken with reserve; also the tolerance of one-quarter of one per cent is more than is usually allowed in achromatic combinations; thus a lens achromatized for the C and F rays shows a deviation for G′ not greater than one-sixth of one per cent. If the larger tolerance were permitted, achromatism with glass would still only extend from approximately 4000 to 8000 A. (The {quartz, lithium-fluoride} combination was first used by B. K. Johnson for a low-power microscope objective—see § 3, Microscopes.)

M. J. Gunn (1936) has proposed objectives made partly of glass and partly of liquid; W. E. Williams (1937) proposes to achromatize lenses of moulded synthetic resin by including a liquid component.

In eyepieces there is little development since the introduction of the wide-angle eyepiece for binoculars a few years ago. For the wide-angle eyepiece it was necessary to abandon the old form of separated field and eye-lenses since, apart from the imperfect correction of astigmatism and distortion,* the eye-point was too

* Spherical aberration and coma are relatively unimportant, as the aperture is small (limited by the pupil of the eye).

close. Better field correction was obtained in the "orthoscopic" eyepiece, where the two lenses were placed close together, the field lens being triple and the eye lens single. When a larger field was required, greater diameters necessitated an arrangement of three separate closely spaced components, of which at least two had to be compounded. Different makers distributed the five elements variously between the three components with almost equal results. Recently Zeiss (1938 a) claim a wide-angle eyepiece having only four elements in two separated components; in another form, Zeiss (1937 a) retain the three-component form, using five or six elements. In none of the eyepieces mentioned is there any attempt at field flatness.

The difficulty of designers in finding anything new in eyepieces is reflected in a recent claim from a patent specification: "... the numerical value of the quotient obtained by subtracting the thickness of [one element] from the algebraic difference between the radii of curvature of this [element] and dividing the remainder by the product of these two radii of curvature is smaller than 1·55 times the reciprocal value of the focal length of the eyepiece"!

§3. MICROSCOPES

With the increasing importance of photomicrography, attention has been directed to the employment of ultra-violet radiation, because resolving power, being inversely proportional to wave-length, is so much greater with light of short wave-length. There are two objections to the use of ordinary visual objectives for photography—the opacity of the glass to ultra-violet radiation and the impossibility of correcting a lens of high aperture for both the visual and photographic regions of the spectrum. If lenses are to be used entirely for photography there is no need for achromatism (focusing is done by trial), so that it is possible to make the objective from a single material; quartz is naturally chosen because of its transparency to ultra-violet. It has long been known that it is possible to correct for spherical aberration a system comprising simple elements of a single material. The minimum number of elements depends on the refractive index (it is four for quartz, for which $n_d = 1{\cdot}544$ for the ordinary ray) (v. Rohr, 1903). With high-aperture systems it is necessary to use more than the minimum. B. K. Johnson (1933) has published data for an objective comprising seven elements of fused quartz, having N.A. = 1·20, for use with light from the Cd line 2140 A.

An interesting development is the reflecting microscope objective. The Mangin mirror, consisting of a meniscus dispersive lens silvered on its convex face to form an aplanatic concave mirror, has been widely used as a condensing reflector (see § 2, Telescopes, for employment as a telescope objective; also § 5, Photographic lenses). B. K. Johnson (1934) has shown that it is capable of being sufficiently well corrected to serve the more exacting conditions of microscopy up to an aperture of N.A. 0·50, which, with ultra-violet light of wave-length 1990 A., gives a resolving power equal to that of a visual objective having a numerical aperture of 1·40. The reflector is almost achromatic, so that it can be used over a wide range of wave-lengths and the focusing done visually; other advantages are that

there is increased contrast in the image owing to the fewer glass-air surfaces, and reflex images with opaque objects are absent; a disadvantage is that a semi-reflecting surface must be used for illuminating even transparent objects.

Another reflecting objective has been constructed by E. H. Linfoot (1938), comprising a spherical reflector and a Schmidt correcting plate. His mirrors have a radius of 3 inches and work at apertures of 0·41 and 0·54, so that the respective diameters are $1\frac{1}{4}$ and $1\frac{3}{8}$ inches; these sizes are ample for the figuring of the correcting plates.

B. K. Johnson (1939) has used the (lithium fluoride, quartz) combination in a doublet microscope-objective having a focal length of 25 mm. and N.A. = 0·20. This is practically apochromatic in the region of both visual and photographic observation, the secondary spectrum in the range 2300 to 6000 A. never exceeding one-tenth of one per cent of the focal length.

A notable advance in microscope optics has been made by the firm of Zeiss with their flat-field objectives (Planachromats). Two methods have been used to reduce the Petzval sum; in the first (Zeiss, 1934) (figure 3) the objective comprises two groups, of which the first (that nearer the object) resembles an ordinary

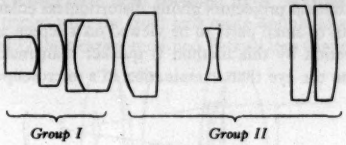

Figure 3.

medium-power microscope objective having a nearly hemispherical front lens followed by a meniscus lens and a compound component. The second group is of low power and comprises a dispersive lens placed between two separated collective components—a kind of Cooke triplet in fact; it has a negative Petzval sum and thus reduces the Petzval sum of the first group.

In the second (Zeiss, 1937 b) (figure 4) the usual hemispherical front lens is

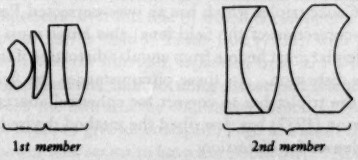

Figure 4.

replaced by a thick meniscus in which the surface nearest the object is dispersive and has its centre of curvature nearly on the object. The rear surface is collective and shallower than the front surface, but by virtue of its thickness the lens is

collective. The contribution of this meniscus to the Petzval sum ($\Sigma 1/f.n$) is thus negative, while its contribution to the power is positive. This front lens is followed by the usual meniscus and compound components, the three forming a member which produces a virtual image at a considerable distance. A second member is placed at a distance of about twice the focal length of the objective from the first, and comprises a similar thick meniscus, facing the image, and a compound component. Each member is separately corrected for spherical aberration and together they are corrected for coma and astigmatism; the combined effects of the two menisci, each corrective to the Petzval sum, result in a considerable approach to the fulfilment of the Petzval condition. The apertures so far made are 0·20 and 0·65. Neither form is likely to be suitable for making objectives of the highest aperture, but they should furnish useful lenses for photomicrography, where the greatest need is felt for flat-field objectives.

§ 4. PROJECTION

(a) *Profile projection*

The increasing accuracy required in modern engineering production has led to the demand for optical projectors giving distortionless enlargement (generally from $\times 20$ to $\times 100$) of small parts to be viewed on a screen and compared with a standard. Inspection by this method is quicker than making measurements and less fatiguing to the eye than examination in a microscope. As a silhouette

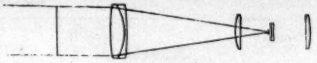

Figure 5.

is projected the object is illuminated by parallel light. The conditions are essentially different from those in kinematography, where a flat object is illuminated by convergent pencils. Lenses for profile projection should accordingly be specially designed. Figure 5 shows a lens designed by H. W. Lee (1934). The collimated light after passing the object falls on a field lens which accordingly forms an image of the light source at its focal plane; in the neighbourhood of this image is placed a Cooke triplet which has an over-corrected Petzval sum to compensate for under-correction of the field lens; this latter must fulfil the tangent condition and the triplet must be free from angular distortion if the projected image is to be free from distortion. In these circumstances the field lens cannot be aplanatic, so that the triplet has to correct for spherical aberration and coma as well. J. S. Anderson (1937) has described the method devised at the N.P.L. for testing projection lenses for distortion.

The success of projection methods in industrial testing has led to its extensive employment in various directions, notably in the saccharimeter and the refractometer. In both cases a straight-line boundary between two fields has to be examined; with the powerful monochromatic light sources now available, this can be done on the projected image.

(b) *Kinematograph projection*

With increase in size of cinema theatres and their screens it has been necessary to increase the light flux through the film and lens; this can only be done at present by using the arc with larger condensing mirrors (the angular aperture is too large to be dealt with by lenses) and greater currents, with the result that the heat flux is also increased (and the mirror, being achromatic and non-absorbent, is much more efficient for collecting heat than the lens), and although this is countered to some extent by artificial cooling at the gate by air-blast, etc., the heat may be enough to affect any balsam used in the projection lens; consequently recent designs have tended to eliminate cemented surfaces. It is undesirable to increase the number of glass-air surfaces, as light scatter is particularly to be avoided in projection systems; consequently it is highly desirable to keep the number of components as small as possible. Usually four is the minimum at present possible for the large-aperture systems required.

Lenses for projecting amateur films are not subjected to such great heat, and cement can safely be employed.

New projection lenses have been patented by A. Warmisham (1938 b) consisting of a Cooke triplet with the front lens made double to reduce aberrations; Kodak (1937 a) modify their earlier modified Petzval lens (figure 6) by introducing

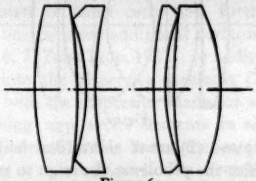

Figure 6.

glass having a refractive index greater than 1·70; Schade (1937 a) has another modified Petzval, in which each of the rear components is a doublet; the aperture is f/1·3; the same author (1937 b) introduces yet another modification by employing two separated components in front, one of which is compound.

A different type of lens is that of Trans-Lux (1935), who place a compound collective component near the film and, separated from it, a compound dispersive component of large diameter, thus forming a reversed telephoto lens with the nodal planes between the film and the collective component, so that a very short focal length can be used, to give an enlarged image at a short throw.

The variable-power lens seems to have lost popularity. At least two members have to move relatively to one another and the focal plane in order to keep the image focused while it changes in size. The movement calls for increased accuracy both in the mechanical and the optical parts, and even so it is extremely difficult to design the system well corrected over a large range of movement. In a recent design H. Chrétien (1934) employs a fixed dispersive lens in front of

two collective lenses, which both move. H. I. Gramatzki (1935), in a variable-power attachment to be placed in front of a corrected lens, employs a telescopic combination capable of being set at two different magnifications. A collective lens is arranged between two separated dispersive lenses each of lower power; when the collective is in juxtaposition to the front it forms a Galilean telescope; when it is moved close up to the rear dispersive, the system is a reversed Galilean telescope.

§ 5. PHOTOGRAPHIC LENSES

Photographic lens design has been influenced chiefly by the needs of: (1) the amateur kinematographer for lenses of still higher aperture, for interior kinematography and for studio work with limited artificial light: the lenses must be of fairly simple design and sell at a moderate price; (2) the miniature photographer for lenses having apertures from f/3·5 to f/2, which must yield sharp negatives as free as possible from scattered light, and with uniform illumination over a fairly wide angle; (3) the aerial photographer and cartographer for lenses of medium aperture f/6 to f/4, yielding evenly exposed negatives of great sharpness over a wide angle, so as to economize in time and film.

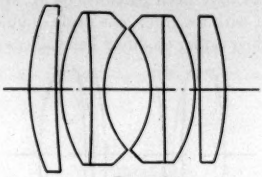

Figure 7.

Lenses in groups (1) and (2) are of short focal length and correspondingly small size. This simplifies the problems of design to some extent, and removes some restrictions as to choice of glass, and the thickness and diameter of the elements can be increased, if desired, to a degree which would make the lens, if of normal focal length, bulky and heavy, and absorb too much light.

Many large-aperture anastigmat lenses are derived from the type which comprises two thick compound dispersive menisci between collective components. Figure 7 (H. W. Lee, 1920) is the prototype. Recent modifications are those of H. W. Lee (1935 b), A. Warmisham and C. G. Wynne (1937) and J. W. Hasselkus and G. A. Richmond (1937). Correction of the Petzval sum follows from the separation of the dispersive from the collective surfaces. This type of lens is capable of giving good definition over a field of 50° to 60° at an aperture of f/2. The aperture can be increased somewhat by reducing the curvatures of the surfaces to reduce the spherical zone, and by increasing the thicknesses of the dispersive menisci to keep the Petzval sum low. Leitz (1936 a) have sought to extend the aperture by a modification, transferring the contact surface in the front meniscus to the front collective lens.

A desirable improvement on the basic type would be to reduce the number of glass-to-air surfaces, and this was done by Zeiss (1929) (figure 8) by removing the rear dispersive meniscus. To replace its correcting power, a surface of equivalent action—that is, a dispersive surface concave to the object—is introduced into the front dispersive meniscus (surface 4). The collective lens 3, 4 must therefore have a lower index of refraction than the dispersive lens 4, 5, an arrangement

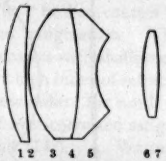

Figure 8.

that is not favourable from the point of view of field correction (Petzval condition); moreover, since it is not possible to get such a large difference of refractive index between two glasses as between glass and air, the surface 4 has to be relatively deep, with large angles of refraction, which is unfavourable from the point of view of zonal spherical aberration and astigmatism. To avoid these drawbacks, and a rather unsatisfactory state of coma correction, further complications have been introduced; thus one or more additional surfaces have been introduced into the rear component 6, 7 (Zeiss Ikon, 1931); or additional cemented surfaces have been introduced into the dispersive meniscus (Zeiss Ikon, 1937a); or additional surfaces into both the dispersive meniscus and the rear component (Zeiss Ikon, 1938), making now seven elements in all. The apertures vary from f/1·6 to f/1·2. They all suffer from the same defect, viz., complete correction for distortion is not achieved. To overcome this last difficulty,

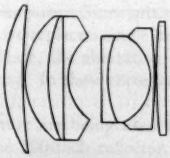

Figure 9.

Zeiss Ikon (1937b) divide the rear component into two (figure 9): it will be seen that there are now just as many glass-to-air surfaces as in the prototype, but the aperture is now f/1·5.

It was remarked, in reference to figure 8, that the smallness of the difference of refractive indices between two glasses led to large angles of refraction at the interface. This has been avoided by A. Warmisham (1936a) by introducing

a glass-to-air contact (figure 10); the correcting power is obtained by a difference in curvature rather than by a difference in index; the advantages are that the curvatures are small, with small angles of refraction, and that it is not now necessary to use a lower refractive index for the collective than for the dispersive element, and this is favourable to obtaining a low Petzval sum; it is to be noted, however,

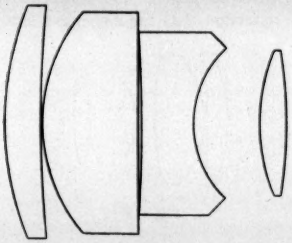

Figure 10.

that in this case achromatism has dictated the use of a highly refractive and dispersive flint; the number of glass-to-air surfaces again amounts to eight. Leitz (1936 b) use a broken contact which is convex to the light, thus further reducing the angle of refraction; the back is triple.

In anastigmat correction, regard must be had to the Petzval sum; if still

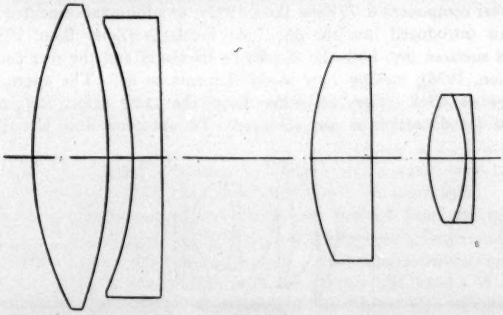

Figure 11.

higher apertures are required, it is reasonable to expect that the field will be smaller till a point is reached at which the designer can ignore this limitation and no longer be under the necessity of employing separated collective and dispersive components. More favourable conditions exist for a large aperture if collective components only are employed, as in microscope objectives where the

aperture is enormous and the field very small. A photographic objective could be made in the same way, but, in order to correct astigmatism, these components must be separated, and this leads to high Petzval sums. A notable example is the Petzval lens itself, which forms the basis of some modern lenses of extra large aperture. Zeiss (1932) have separated the elements of the front component and added an extra collective element to the rear component, which is moved nearer to the focal plane. The "broken contact" gives greater correction for both spherical aberration and astigmatism. The aperture reached is f/0·85 (figure 11). In order to achromatize with shallow curves, the dispersive elements are of extra dense flint, having a high index of refraction, and one of the collectives of borosilicate crown of very low index: the result is a very high Petzval sum, so that a considerable amount of over-corrected astigmatism has to be introduced: the field is consequently small (14°). A. Warmisham (1936 b) has a lens of aperture f/0·9 (figure 12); it is a modification of his own modified Petzval lens

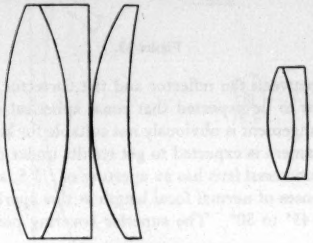

Figure 12.

(Warmisham, 1921), in which the front is split into two, a well-known device which has been noticed more than once. Since spherical aberration is proportional to the cube of the aperture, if we replace a single lens by two, each of double the focal length and half the aperture, the aberration of the two will be $2 \times 2^{-3} = \frac{1}{4}$ the aberration of the single lens. In these three lenses the distortion coefficient is large.

Lenses of still larger aperture are being used as part of the spectrographic equipment associated with the 200-inch reflector, but as these are practically enlarged microscope-objectives, there is no need to discuss their design. Zeiss (1936 c) (figure 13) use a reflecting surface in a system having an aperture of f/0·5. Here again the Mangin type of reflector is employed, and with it, in front, is a correcting component comprising separated collective and dispersive elements, having a weakly dispersive effect altogether, which correct astigmatism, so that the lens covers a field of 15°. The advantages of a reflecting surface over a refracting one of the same radius are: (1) the power of the reflecting surface is $2n/r$ and that of the refracting surface $(n-1)/r$, where n is the refractive index of

the medium in which the ray is travelling before it meets the surface; thus for a given power the reflector has a radius 5 to 6 times as large; (2) the Petzval contribution of the reflecting surface is $2/n\nu$; of the refracting surface $(n-1)/n\nu$. The respective values of the ratio (Petzval sum/power) are $1/n$ and $1/n^2$; so in this respect the reflecting surface is n times as favourable; (3) the spherical-aberration coefficients in the two cases are $\frac{1}{4}$ and $(n/(n-1))^2$. If $n=1{\cdot}5$ these are $\frac{1}{4}$ and 9, so that the reflector is 36 times as favourable. This objective, like the Schmidt,

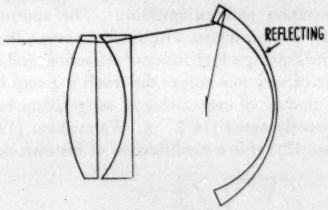

Figure 13.

has its focal plane between the reflector and the corrector, but in this case the field is flat; it is not to be expected that zonal spherical aberration is entirely removed. The arrangement is obviously not suitable for kinematography.

The miniature camera is expected to get results under difficult conditions of lighting so that the universal lens has an aperture of f/3·5, and covers an angular field of 60°, while lenses of normal focal length at this aperture are not expected to cover more than 45° to 50°. The superior covering power of the miniature

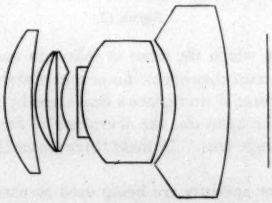

Figure 14.

camera lens is possible on account of its small size, but with the larger angle there is greater decrease of light at the margins of the photograph, partly because the illumination falls off according to the fourth power of the cosine of the angle (hence to 56 per cent at 30° from the axis), but more from the cut-off by the rims of the lenses, and this is particularly marked with lenses of large aperture (Chilton, 1937). Designers have sought to remedy this last by increasing the diameters of the outer lenses. In the Zeiss Biogon, f/2·8 (Zeiss, 1935 b) (figure 14), the rear

lenses only are enlarged, and as these are inside the camera, marginal illumination is increased without increase of external size. At the same time the rear components are made so thick that the back surface is quite close to the focal plane. Increasing the marginal aperture increases the effects of coma and astigmatism, which, therefore, have to be much better corrected than usual; this explains the rather complicated structure. The basic form is that of figure 8, with a triple back lens which has been given a split contact. Leitz (1939), in their Summitar, have preferred to enlarge the front lens in order to increase marginal illumination.

The ever-recurring problem of reducing the number of glass-to-air surfaces is of particular importance in miniature camera lenses. H. W. Lee (1936) has

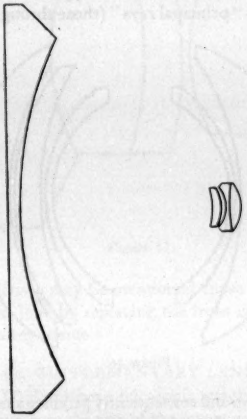

Figure 15.

reduced the number to six while maintaining the field at 45° at an aperture of f/2. Outwardly the form is that of a Cooke triplet, but each collective lens is a doublet and the dispersive is a triplet; the contact surfaces play a considerable part in the correction. There are here three collective elements and four dispersive elements in all, and this makes for easy chromatic correction with shallow curves; as the collectives are made of glass of high refractive index and the dispersives of glass of relatively low index, good correction of the Petzval sum is possible. A six-element Cooke type lens has also been introduced by Leitz (1931) as the Hektor.

Most of the photographic lenses dealt with have been of high aperture; there are some in which attention is directed to the field. Zeiss (1935 a) (figure 15) make a lens embracing the whole hemisphere at an aperture of f/8; this field is

useful to meteorologists in photographing cloud formations. As previously employed by R. Hill (1923), a large negative lens first reduces the 180° field to about 100° and passes it on to a rear component of collec ive action, which takes the well-known form, in this case, of the rear member of a Goerz Dagor lens with "split contact." It need hardly be said that distortion is not corrected.

J. W. Hasselkuss and G. A. Richmond (1936), modify a well-known type of wide-angle lens which comprises two thin strongly curved meniscus dispersive lenses between two meniscus collective lenses, covering a field of about 100° at an aperture of f/8; they replace the rear collective lens by two meniscus collective lenses (figure 16), and are able thereby to increase the aperture to f/5·6, which makes it suitable for air survey work. In this type of lens it is necessary to make the curves deep so that the "principal rays" (those through the centre of the stop)

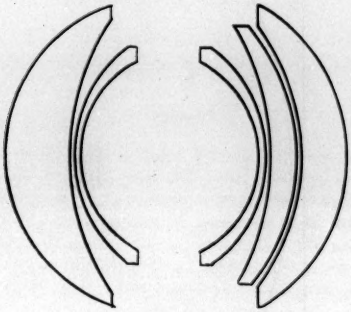

Figure 16.

have small angles of refraction and consequently produce small amounts of astigmatism; to keep the Petzval sum from being over-corrected the lenses must be thin.

One of the troubles encountered in the use of wide-angle lenses is the possibility of flare spots, ghosts and other spurious light effects, which are especially prominent when the iris is stopped down, and this is most generally the case with wide-angle lenses. These flares arise chiefly from internal reflection in the components behind the diaphragm, because the iris, when stopped down, cuts off most of the effect from the front member. Two members are usually employed, one on either side of the diaphragm, to correct distortion (very important in wide-angle lenses, as this varies as the cube of the field); a collective lens in front of the diaphragm inevitably has pincushion distortion, while one behind the diaphragm has barrel distortion. Zeiss (1936 d) (figure 17) have sought to correct the distortion of a lens of Goerz Dagor type in front of the stop by placing in front of it a meniscus of weak dispersive power having its concave surface turned towards the object; a small amount of under-corrected astigmatism in

this lens, by virtue of its relatively large distance from the stop, introduces the necessary correcting barrel-shaped distortion. All the elements of this objective are in front of the diaphragm, and flare spots, etc., should consequently be absent or, at any rate, much reduced as compared with a symmetrical type of lens.

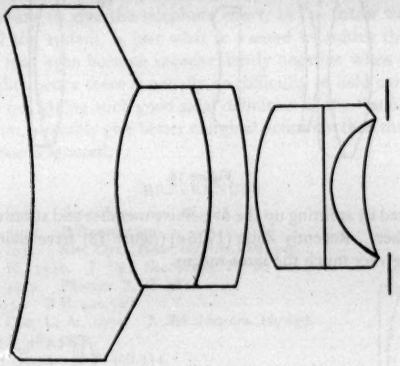

Figure 17.

Among other new lenses may be mentioned those of A. W. Tronnier (1936), who modifies the Tessar lens by replacing the front component (which is single in the Tessar) by two components.

§ 6. SUPPLEMENTARY LENSES

J. A. Schneider (1937) describes a telescopic attachment to be used in front of a photographic lens to increase the size of the image without altering its position, Kodak (1937 b) an attachment to be placed behind a photographic lens to shorten the focal length and increase its aperture, so that an f/2 lens becomes one of f/0·813, with a field of 9°. Thus the focal length is reduced $2\frac{1}{2}$ times and the field (assuming the original lens to have a normal field of about 40°) reduced $4\frac{1}{2}$ times; the dimensions of the field covered will thus be reduced eleven times. The system comprises a meniscus collective lens and a compound collective lens which is so thick that it reaches nearly to the focal plane.

§ 7. TELEPHOTO LENSES

The telephoto was introduced as a compact form of variable long-focus lens. With the improved standard of definition introduced by the anastigmat, the variable-power telephoto was found wanting and has more or less disappeared, its place being taken by the fixed-focus lens of moderate focal-length having anastigmat definition. Since the construction requires a collective member

in front of the diaphragm and a dispersive member behind it, both of which have distortion of the same sign, it was considered that pincushion distortion was inevitable in a telephoto. However, H. W. Lee (1923) had shown that distortion

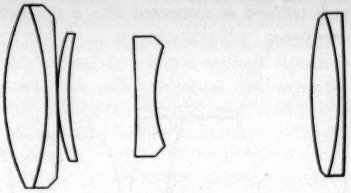

Figure 18.

could be corrected by splitting up the dispersive member and subdividing the front collective member. Recently Zeiss (1936e) (figure 18) have claimed to correct distortion, using very much the same means.

Figure 19.

Telephoto lenses nowadays find their greatest use in kinematography, and the aperture has had to be increased. A. Warmisham (1937) (figure 19) has

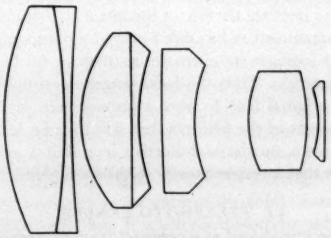

Figure 20.

reached an aperture of f/1·65 by making the collective member of two separated collective doublets. The I.G. für Farbenindustrie (1937) (figure 20) make the

collective member of two collective doublets and a single dispersive lens, and the dispersive member is a separated doublet; the aperture is f/1·4. Wood (1934) has an arrangement similar to Warmisham's but with the last lens simple; the aperture is f/1·6. The large separation between the collective and dispersive members, necessary to give the telephoto effect, i.e. to throw the nodal planes well in front of the system, is just what is wanted to reduce the Petzval sum, which, in fact, may even become inconveniently negative when large telephoto effects are sought; hence there is usually no difficulty in field correction, so that these lenses, if not giving such good axial definition as the "straight" lenses of the same aperture, probably give better marginal definition than those in which the Petzval correction is ignored.

REFERENCES

ANDERSON, J. S., 1937. *Proc. Phys. Soc.* **49**, 112.
ASKANIA WERKE, 1934. B.P. 426,539.
BRICOUT, M. P., 1931. *Rev. Opt.* (*théor. instrum.*) **10**, 345.
CARTWRIGHT, C. H., 1939. *J. Opt. Soc. Amer.* **29**, 29 and 350.
CHILTON, L. V., 1937. *Photogr J.* 77, 381.
CHRÉTIEN, H., 1934. B P. 440,397.
COX, H. W and COX, L. A , 1939. *J. Sci. Instrum.* **16**, 257.
DJIAN, 1936. B P. 489,381.
GRAMATZKI, H. I., 1935. B P. 449,434.
GUNN, M. J , 1936. B.P. 468,227.
HASSELKUS, J. W. and RICHMOND, G. A., 1936. B.P. 472,191.
HASSELKUS, J. W. and RICHMOND, G. A , 1937. B.P. 507,590.
HILL, R., 1923. B.P. 225,398.
HILL, R., 1926. *Proc. Opt. Conv.* 878.
I.G. FÜR FARBENINDUSTRIE, 1937. B.P. 513,452.
JOHNSON, B K., 1933. *J. Sci. Instrum.* **10**, 174.
JOHNSON, B. K., 1934. *J. Sci Instrum.* **11**, 384.
JOHNSON, B. K., 1939. *Nature, Lond.*, Jan. 31.
KODAK, LTD., 1934. B.P. 461,259.
KODAK, LTD., 1936. B.P. 503,673.
KODAK, LTD., 1937 a. B.P. 517,271.
KODAK, LTD., 1937 b. B.P. 517,499.
LEE, H. W., 1920. B.P. 157,040.
LEE, H. W., 1923. B.P. 222,709.
LEE, H. W., 1934. B.P. 444,350.
LEE, H. W., 1935 a. B.P.s 435,149 and 435,947.
LEE, H. W., 1935 b. B.P. 461,304.
LEE, H. W., 1936. B.P. 477,448.
LEE, H. W., 1937. B.P. 503,589.
LEE, H. W., 1939. *Proc. Phys. Soc.* **51**, 786.
LEITZ, 1931. B.P. 382,246.
LEITZ, E., 1936 a. B.P. 480,643.
LEITZ, 1936 b. B.P. 480,641.
LEITZ, 1939. *Kleinfilm Foto*, **9**, 10.
LINFOOT, E. H., 1938. *J. Sci. Instrum.* **15**, 405.
MCLEOD, J. H. and ALTMANN, F. E., 1938. *J. Soc. Mot. Pict. Engrs*, **31**, 40.
MERTÉ, W., 1935. U.S.P. 2,063,178.
v. ROHR, M., 1903. *Formation of Images in Optical Instruments* (Eng. trans.), p. 237.

Ross. F. E., 1935. *Astrophys. J.* **81**, 156.
Schade, 1937 a. B.P. 517,503.
Schade, 1937 b. B.P. 517,504.
Schalz, H. R , 1932. *Z. Phys.* **79**, 843.
Schmidt, B., 1931. *Zentztg. Opt. Mech.* **52**, 79.
Schneider, J. A., 1937. B.P. 479,187.
Stockbarger, D. C., 1937. *J. Opt. Soc. Amer.* **27**, 416.
Trans-Lux, 1935. B.P. 453,171.
Tronnier, A., 1936. B.P.s 476,348/9 and 476,349.
Waland, R. C., 1938. *J. Sci. Instrum.* **15**, 339 ; *J. Brit. Astr. Assoc.* **48**, 308.
Warmisham, A., 1921. B.P. 180,949.
Warmisham, A., 1936 a. B.P. 477,424.
Warmisham, A., 1936 b. B.P. 485,096.
Warmisham, A., 1937. B.P. 492,311 ; Habell, K. J., 1938. *J. Brit. Kin. Soc.* **1**, 2
Warmisham, A., 1938 a. B.P. 514,441.
Warmisham, A., 1938 b. B.P. 517,386.
Warmisham, A. and Wynne, C. G., 1937. B.P. 507,184.
Wild, 1937. B.P. 476,147.
Williams, W. E., 1937. B.P. 490,381.
Wood, 1934. U.S.P. 2,076,190.
Zeiss, 1929. B P. 350,323.
Zeiss, 1932. U.S.P. 1,967,836.
Zeiss, 1934. B.P. 438,860
Zeiss, 1935 a. B.P. 461,554.
Zeiss, 1935 b. B.P. 459,739.
Zeiss, 1936 a. B.P. 478,494.
Zeiss, 1936 b. B.P. 476,153.
Zeiss, 1936 c. B.P. 482,874.
Zeiss, 1936 d. B.P. 487,712.
Zeiss, 1936 e. B.P. 493,650.
Zeiss, 1937 a. B.P. 509,990.
Zeiss, 1937 b. B.P. 506,321.
Zeiss, 1938. B.P. 509,585.
Zeiss Ikon, 1931. B.P. 383,591.
Zeiss Ikon, 1937 a. B.P. 483,802.
Zeiss Ikon, 1937 b. B.P. 497,550.
Zeiss Ikon, 1938. B.P. 499,290.

Recent articles dealing generally with the development of lens optics include:

Kingslake, R., 1940. *J. Soc. Mot. Pict. Engrs*, **34**, 76.
Kingslake, R., 1940. *J. Appl. Phys.* **11**, 56.
Mees, C. E. K., 1939. *Photogr. J.* **79**, 225.
Taylor, W. and Lee, H. W., 1935. *Proc. Phys. Soc.* **47**, 502.

ADDENDUM

Continued interest in the Schmidt telescope is shown by recent articles. A. Conder (1940) and F. A. Lucy (1940) have written on the calculation of the form of the correcting plate; J. G. Baker (1940 a) has suggested a solid-glass Schmidt camera for nebular spectrography. The aperture is f/0·3 and the field 7 degrees. The complete 5th-order aberrations have been worked out and the equation of the meridian section of the corrector given to 10th-order terms. In a further article (1940 b) the same author describes a family of flat-field cameras,

equivalent in performance to the Schmidt, having correcting plate, primary aspheric concave mirror (without points of inflection) and secondary concave mirror.

A number of new photographic systems take advantage of the new dense barium flints now available, which have a high refractive index similar to that of ordinary extra dense flints but with very much less dispersion. The English manufacturers attain a refractive index of 1·653 with $\nu = 46{\cdot}2$; the Germans have pushed these properties further and reached a refractive index of 1·672 with $\nu = 47$. The glass contains a high proportion of barium and is liable to tarnish badly, so the wisdom of employing it unprotected is doubtful. Perhaps it is generally a case of paper design running ahead of manufacture. Usually the new material has been used to improve an earlier type of design, but a more interesting application is to be noted. The Cooke lens consists of three elements separated by air, thus involving six glass to-air surfaces. The first two comprise a collective of high refractive index separated by an air space from a dispersive element. By employing a glass of still higher index it is possible to get a sufficient difference of index with a low refracting glass as the second medium instead of air. The first two air-spaced elements of the Cooke lens are thus replaced by a cemented triplet, and the total number of glass-to-air surfaces reduced to four. Zeiss (1939) have introduced this in Germany, and both Zeiss and Tronnier (1938) have applied the same principle to the Tessar lens.

REFERENCES

Baker, J. G., 1940 a. *Proc. Amer. Phil. Soc.* **82**, 323.
Baker, J. G., 1940 b. *Proc. Amer. Phil. Soc.* **83**, 339.
Conder, A., 1940. *C.R. Acad. Sci., Paris*, **210**, 327.
Lucy, F. A., 1940. *J. Opt. Soc. Amer.* **30**, 237.
Tronnier, A. W., 1938. B.P. 517,946.
Zeiss, 1939. G.P. 652,882; Huber, S., *Photogr. Ind.* 1076 and 1092.

THE ABSORPTION OF LIGHT IN INTERSTELLAR SPACE

By A. HUNTER, Ph.D., F.R.A.S.,
Royal Observatory, Greenwich

§1. INTRODUCTION

It comes as something of a shock today to find in a book on astronomy (one still recognized everywhere as a standard textbook on the subject) a section entitled "Evidence against the absorption of light in space"; and then to learn from its title-page that the volume was first published as recently as 1927. In the past few years astronomical journals have devoted a large and increasing proportion of their space to studies of interstellar absorption—studies which attack the subject from many different standpoints. This review deals with the spectroscopic line of approach.

The question whether matter exists between the stars is not a new one: Barnard's "dark markings" in the Milky Way, now known to be due to absorption, were a source of lively controversy in the opening years of this century. Evidence for the existence of such matter accumulated gradually: the presence of interstellar dust was established by direct photography, by the analysis of star-counts, and by the discovery of reflection nebulae and of the reddening of starlight; that of interstellar gas by the recognition of the lines of H, He and O in the spectra of emission nebulae, and by the discovery of "stationary" lines in stellar spectra, showing that Ca, Na and other elements were widely distributed in space. But the importance of the subject was not fully appreciated until more recently, when the feeling took root that the effects of interstellar absorption on cosmology were likely to be too large for continued neglect. Finally, with the recognition of the spiral nebulae as external galaxies on the same model as our own, it became clear that the pockets of absorption seen in these other universes were to be compared with our own dark nebulae, whilst the dark band stretching across any spiral seen edge-on must have its counterpart here in widespread absorption in the plane of the galaxy.

The importance of interstellar light absorption is threefold. Firstly, the pure astronomer interested in the structure of the universe needs the assurance that his stellar distances either are not affected by absorption, or can be corrected to allow for it, if necessary. He must, in fact, know with reasonable certainty whether his distant stars look faint because they *are* faint or because their light is partly lost on its way to us. Secondly, the astrophysicist taking the evolution of the universe as his study wants to know whether any appreciable part of the

mass of the galaxy remains uncondensed into stars, and if so, what are the physical and chemical constitutions of this interstellar matter. Finally, the physicist cannot fail to be interested in the behaviour of matter under interstellar conditions; matter at unattainably low pressures, at temperatures he can reach only with difficulty, and in radiation fields which he cannot hope to reproduce. And as is so often the case in astronomy, exploration of the subject from all three points of view involves enlisting the aid of spectroscopy.

The principle on which the spectroscopist works in investigating interstellar absorption is quite simple. The light we receive from a distant star has been on its journey some few centuries. The tenths of a second which it spends crossing the stellar atmosphere and the thousandths crossing the terrestrial atmosphere leave their indelible marks on the spectrum, but these marks are characteristic, easily recognized, and by now well known. The remaining absorption must have arisen in the intervening space. It shows itself in two ways:—

(*a*) by absorption at discrete wave-lengths (by atoms and molecules, neutral or ionized), and

(*b*) by continuous absorption, selective or non-selective (by electrons, atoms, molecules and molecular aggregations up to dust particles and meteorites).

In neither case is this absorption the most striking feature of the spectra of even the most distant stars, whilst generally speaking no interstellar absorption at all can be detected in the light from our nearer neighbours in space. Taking into account the enormous light-paths involved, this fact shows that any interstellar matter must be extremely tenuous; nevertheless, just because the paths are so long for those stars in the outer regions of space which are our galactic milestones, the absorption, small as it is per unit distance, takes on importance to astronomers.

§2. LINE ABSORPTION

What can the spectroscopist expect to deduce from a study of interstellar line-absorption in stellar spectra? First of all, he recognizes the inherent observational difficulties. This absorption is always accompanied by atmospheric absorption (stellar or terrestrial), which is usually as strong or stronger. Unless the particles in interstellar space are of a kind not represented in stellar or planetary atmospheres, difficulties will arise in disentangling one absorption from the others. And unfortunately for the present purpose, Nature has been methodical (or perhaps merely unimaginative!) in her distribution of cosmic material: the elements abundant in the stars and planets (or rather, those well represented in stellar and planetary spectra) are also those common in interstellar space. Luckily there are means of separating the three absorptions. In the spectra of stars with variable radial velocity (e.g. double stars whose orbital planes lie in or nearly in the line of sight), the Doppler effect gives all the lines arising in the stellar atmosphere a periodic displacement which is exhibited by neither terrestrial nor interstellar lines. The terrestrial lines cause no difficulty: it is not for nothing that the solar spectrum has been subjected to intensive

scrutiny for decades. The remaining lines (called " stationary " or " detached " lines) are of interstellar origin; the first examples were discovered in 1904 in the spectrum of δ Orionis. They were the H and K lines of Ca II at 3968, 3933 A. These lines were also noticed to be very much sharper than the average stellar lines, a circumstance which made their recognition in other non-binary stars all the easier. This feature is a natural consequence of the low pressure at which the lines originate; a pressure considerably lower than that in the atmospheres even of supergiant stars showing the " *c* " spectral characteristic (extreme sharpness). Finally, stellar absorption may be distinguished from interstellar if the excitation conditions in the stellar atmosphere are widely different from those in interstellar space. For instance, the interstellar H and K lines appear without contamination by stellar Ca II lines in the spectra of the hottest stars (types O, B_0, B_1, B_2, B_3), though calcium is, of course, present in their atmospheres. Fortunately these stars are intrinsically the most luminous in the sky; consequently they can be seen at great distances and would be expected to show any absorption which is present.

The observational difficulties mentioned above have not prevented the recognition of many more elements in interstellar space. The presence of Na was established in 1919, when the D lines were found; since then the second doublet of the principal series at 3302, 3303 A. has been added to the list of interstellar lines. The first doublet of the P series of K I, or rather the one member of it at 7699 A. which has escaped obliteration by terrestrial absorption; the Ca I line at 4226 A.; and Ti II lines at 3073, 3229, 3242, 3384 A. are of more recent recognition (Dunham, 1937). Other elements such as Al, Fe, Sr and Ba might be expected in interstellar space since they are certainly present in the stars themselves; moreover, they have astronomically observable lines. That they have not been found may be a matter of low abundance either of the elements themselves or (which is more likely) of their observable ions.

On the other hand, some lines of interstellar origin have been observed but not yet certainly identified. Lines of a group in the violet at 3875, 3934, 3957, 4233, 4300 A. are all very sharp. Suggestions have been made (Dunham, 1939) that they represent upward transitions from metastable states from which the downward transitions are strongly forbidden. However, an hypothesis due to McKellar (1940) that they are molecular in origin seems more likely in view of recent developments. McKellar has provisionally identified the lines at 3875, 3934 and 4300 A. with lines arising from the (0, 0) bands of CN, NaH and CH respectively. Striking confirmation of this view has come very recently from Mt. Wilson (Adams, 1940), where some fainter lines of CH predicted by McKellar have been found on a high-dispersion spectrogram of ζ Ophiuchi at the predicted wave-lengths and with the predicted intensities. The evidence for the presence of CH molecules in interstellar space thus seems conclusive. In the yellow and red regions of the spectrum, diffuse lines at 5780, 5796, 6284, 6614 A. (Merrill, 1934) have also been attributed to molecules (Russell, 1935), though

reported coincidences with CO_2 vibration-rotation band-heads for the two red lines (Swings, 1937) may be coincidences in more senses than one. According to another interesting hypothesis (Merrill, 1937) they are due to quasi-line absorption by solid particles.

So far, we have dealt only with the qualitative aspect of the subject. It is a matter of interest for physicist and astronomer alike to know the actual density of atoms of Na, Ca, K, Ti, etc., in interstellar space. From the ionization equation

$$h\nu + \mathrm{Ca} = \mathrm{Ca}^+ + e$$

can be derived the free-electron density by substituting a reasonable photon density (calculated from the known space-density of stars of different temperatures) and the observed ionization ratio for interstellar Ca, this latter being derived from the relative intensities of Ca I 4226 and Ca II 3933. The density ratio of any two consecutive ions of any other element can then be computed from the appropriate ionization equation, using this density of free electrons. Summation then gives the actual abundances of the various elements when one is known. The electron density found by different authors varies from 3 cm.$^{-3}$ (Strömgren, 1939) to 30 cm.$^{-3}$ (Struve, 1939), a not unlikely value being 10 cm.$^{-3}$ (Dunham, 1939). In any case the figure is far higher than the number of all the known metallic positive ions combined. The only likely source for such a high concentration of free electrons is the ionization of H atoms. This in turn postulates a high interstellar abundance of hydrogen, a deduction which is supported by the discovery (Struve and Elvey, 1938) that in certain widely-extended areas of the Milky Way the line Hα appears faintly in emission. The light is no doubt a fluorescent emission from H atoms excited by the integrated ultra-violet radiation from the O and B stars, which concentrate close to the Milky Way. This discovery of interstellar hydrogen gives confidence in the present view of interstellar space as deduced from line emission and absorption (Dunham, 1939): in 10 cubic metres these are believed to be of the order of 10^7 free electrons, 10^7 protons, 10^6 hydrogen atoms, 10^3 sodium atoms (mostly as Na^+), 150 potassium atoms (mostly as K^{++}), 60 calcium atoms (mostly as Ca^{++}) and 1 titanium atom (Ti^{++} or Ti^{+++}). The quantities of the newly discovered molecules are not yet known, but CH is likely to prove abundant.

Although such figures as these must be regarded as provisional, and will no doubt be amended by future work, observational and theoretical, it is already apparent that our previous ideas of the amount of atomic matter in interstellar space must be revolutionized. Until quite recently it was firmly believed that apart from the clouds of dust, comets and meteors, which contribute negligibly to the total mass of the galaxy, practically all the material of the universe was condensed into stars. Such abundances as those quoted above, however, indicate that free atoms and electrons are present between the stars in such great numbers that the combined mass of this almost invisible material must be at least comparable with that of the stars. The reaction on theories of stellar evolution is yet to come, but its direction can be foreseen. The abundance *ratios* above

were not unsuspected; and the current theory that interstellar atoms have been blown away from the hotter stars by radiation pressure (much as we see, day by day, eruptive prominences blown away from the solar surface) was not incompatible with such ratios, since interstellar material is not totally different from stellar. But that the large quantities indicated by this recent work were ejected in this way seems less likely than that the interstellar matter is primaeval; in fact, that it represents the debris left over when the stars condensed from the primitive all-pervading nebula.

This hypothesis is supported by what we know of the spatial distribution of the matter. It is virtually everywhere in the galactic system. The intensities of the various lines correlate well with one another, indicating that the material is well mixed. Moreover, the intensities correlate well with distance; so well that, though local irregularities do exist (Beals, 1936; Gleissberg, 1937), "the intensities of the interstellar lines may be considered as giving the most reliable known method of obtaining the individual distances .... of the stars in which they appear" (Plaskett and Pearce, 1933). Finally, exact wave-length measurements support the hypothesis that the absorbing atoms share in the general galactic rotation. The study of interstellar absorption lines therefore provides us with a general picture of our universe in which the stars are seen to be embedded in a rarefied gas, homogeneous to a first approximation, and extending out at least to the most distant members of the galaxy—perhaps thinning out there as do the stars. The gas consists of electrons, protons, atoms of H, O, Ca, Na and Ti (and possibly of other less easily identified elements) and molecules, certainly CH and probably others too. It is stationary with respect to the mean of the stars, though local currents do exist, and it contributes by no means negligibly to the total mass of the galaxy.

§3. CONTINUOUS ABSORPTION

The line absorption which we have so far considered abstracts very little energy from the continuous spectra of stars. For that reason it is not regarded with so much dismay by the astronomer interested in the structure of the universe as is continuous absorption, for which evidence has become overwhelming during the past decade.

It is at once one of the most important and one of the most difficult problems of astronomy to deduce from the apparent distribution of the stars in the sky their true distribution in space. The many hypotheses current at various times as to the structure of the universe have all sprung from different solutions of this problem; and have represented the true structure well or ill according as the respective solutions were based on well- or ill-founded estimates of stellar distances. But accurate distances are not easy to come by. By trigonometric means we can reach out only a few hundred light-years, which is probably little more than a thousandth part of the maximum extent of the galaxy. Beyond that we must rely on indirect means. The fundamental difficulty is to know whether a star

which appears faint is really of low luminosity or whether it is an intrinsically bright object at a great distance. Fortunately for the stellar statistician, independent evidence is available as to the true luminosities of many kinds of star. Frequently this evidence is spectroscopic in origin: for instance, nearly all stars of early spectral types (O, B, and A) are also intrinsically bright; and stars of late type can often be assigned a definite luminosity by using delicate spectral criteria which are functions of the density of the star's atmosphere, and thence of its surface gravity and mass. Another powerful though limited method is by using the known correlation between the luminosities of certain variable stars and their periods of light-variation.

Such evidence as to stellar luminosities leads to what may be called photometric distances of the stars: that is, to distances derived by comparing their true and apparent brightnesses. In terms of stellar magnitudes (of which, it will be recalled, the unit is a light-ratio of $\sqrt[5]{100}$, since first-magnitude stars are a hundred times brighter than sixth), the difference between the apparent brightness m of a star and its brightness M ("absolute magnitude") at the standard distance of 10 parsecs may be derived in terms of its distance r as follows. Let the luminous flux received from the star at its true distance r parsecs be l_r. Then at the standard distance the flux l_{10} received would be by the law of inverse squares

$$l_{10} = l_r \cdot r^2/10^2. \qquad \ldots\ldots(1)$$

By definition $$l_r/l_{10} = (\sqrt[5]{100})^{M-m}$$

or $$\log (l_r/l_{10}) = 0{\cdot}4\,(M-m). \qquad \ldots\ldots(2)$$

Combining (1) and (2) we have

$$\log r = 0{\cdot}2(m-M)+1, \qquad \ldots\ldots(3)$$

the right-hand member of which is an observational quantity. It will be noticed that values of r obtained from (3) will be correct only if (1) is valid. If light is absorbed in interstellar space, estimates of cosmic distances will be too great, since dimming of a star's light by absorption will be falsely attributed to inverse-square dimming by distance. Or, looking at the matter in another way, the values of m in (3) will be increased by absorption, leading to an over estimate of r. The whole scale of cosmic distances thus depends on the validity of (1), or at least on the possibility of accurately correcting the relation for any divergence from observational fact. Clearly if anything definite is to be learned about the structure of the stellar universe, the question of absorption must be settled; and in settling it spectroscopy can play an important, if not this time a crucial, part. Doubt no longer exists that absorbing matter in the form of finely-divided solids is present between the stars; the diffuse nebulae, dark and bright, are plain to see, and indirect evidence in plenty is available in the "zone of avoidance", where extra-galactic nebulae are obscured by absorption, and in the reddening of the light of distant stars. It is in this last field that spectroscopy and spectrophotometry are particularly useful.

If the particles responsible for the absorption permeate interstellar space and are between certain wide limits of size, the absorption will increase with decreasing wave-length and will redden the starlight more and more with increasing distance. The observed colour of a star, however, is primarily a function of its photospheric temperature, and thence of its spectral type. The problem of deciding whether the presence of interstellar absorption makes the colour secondarily a function of distance is much more complicated than it seems at first sight. The colour observed depends on the method of observing it, whether it is by measuring the colour index (photographic or photoelectric), the spectrophotometric gradient, or by tracing, wave-length by wave-length, the spectral distribution curve; and the relations between the several parameters are complicated (Atkinson *et al.*, 1940). For all stars but the closest ones the distance is to some extent uncertain because of the absorption, the very quantity which is the subject of the investigation. Finally, stars of the same spectral type at the same distance in the same direction of space may differ in colour if their luminosities differ. This is particularly serious because of the unavoidable observational selection in any programme of stellar colour measurements; the stars will not differ widely amongst themselves in apparent magnitude, so that distance will correlate at least partially with luminosity, and part of any effect apparently due to distance may well be due to an absolute magnitude effect.

However, it appears (Hunter and Martin, 1940 a) that for the hottest stars the dependence of colour on luminosity is small; and luckily these are the most luminous and most distant stars and are thus particularly suited for study in investigating interstellar space. Normal colours can therefore be assumed for each spectral sub-type, and the reddenings, which represent divergences from these normal colours, can be analysed with respect to galactic latitude, galactic longitude, distance, etc. Several investigations of the colours of B stars have proceeded along these lines; the most recent (Stebbins *et al.*, 1939; Hunter and Martin, 1940 b) agree that the observed reddening is confined fairly closely to the Milky Way, stars more than 30° away from the galactic circle being rarely heavily reddened; that it is only roughly proportional to distance at any galactic longitude; and that there is a marked variation of the reddening coefficient with galactic longitude. Stebbins and his collaborators (1940) conclude that the effect of the interstellar material is too irregular to be represented by a mean coefficient of selective (or indeed of total) absorption. Hunter and Martin (1940 b), however, find that mean coefficients at the various longitudes can satisfactorily represent the colours of over 90 per cent of their stars, and that much of the remaining abnormal redness is associated with known obscuring clouds. They therefore suggest with Seares (1940) that there exists a diffuse absorbing stratum, confined fairly closely to the galactic plane and extending at least 1500 parsecs from the sun; and that embedded in this are local condensations which produce abnormally heavy reddenings in certain regions and which may be associated with the obscuring clouds outlining the zone of avoidance of the extra-galactic

nebulae. The order of magnitude of the selective absorption in the plane of the galaxy is a quarter of a magnitude per kiloparsec, if the measurements are made in "photographic" and "visual" light. As mentioned above, however, large variations exist between the different longitudes, and a mean value round the galactic equator has only a limited significance. It is not surprising that the highest values occur towards the galactic centre, and the lowest towards the anticentre.

In order to make full use of the spectroscopic data, we must, if possible, proceed from the observed reddening of a star to the total absorption its light has suffered. Measurements of colour excess, whether in terms of colour indices or of spectrophotometric gradients, tell us the difference of absorption between two regions of the spectrum. For correcting the cosmic scale of distance implied by (3) we must know the total absorption for the light in which m and M are measured. Spectrophotometric observations must be invoked here. The way in which the absorption depends upon wave-length is a function of the size of the scattering particles. The various laws which might be expected range from the Rayleigh λ^{-4} law, highly selective scattering by atoms, molecules, or particles much smaller than the wave-length of the incident light, to the λ^0 law, or non-selective blocking by large particles. Spectrophotometric comparisons of distant stars showing heavy reddening with close unreddened stars of the same spectral type (Struve *et al.* 1934; Stebbins and Whitford, 1936) indicate a type of absorption intermediate between these extremes: probably proportional to λ^{-1} (Strohmeier, 1939; Stebbins *et al.*, 1939), or in some cases to a power series in λ^{-1} (Barbier and Chalonge, 1940).

Let us assume the λ^{-1} law to hold thus:—

$$A_\lambda = A_1/\lambda, \qquad \ldots\ldots(4)$$

where A_λ is the absorption in magnitudes at wave-length λ in A., and A_1 is a constant for the star concerned. If the true magnitudes of the star at effective wave-lengths λ_r and λ_b in the red and blue regions of the spectrum respectively are m_r, m_b, the corresponding values observed in the presence of absorption will be $m_r + A_1/\lambda_r$, $m_b + A_1/\lambda_b$. The true colour-index of the star,

$$C = m_b - m_r,$$

will be modified by the absorption to

$$C' = m_b + A_1/\lambda_b - (m_r + A_1/\lambda_r),$$

and the colour excess

$$\begin{aligned} E &= C' - C \\ &= A_1(1/\lambda_b - 1/\lambda_r) \\ &= 4{\cdot}55 \times 10^{-5} A_1 \qquad \ldots\ldots(5) \end{aligned}$$

for measurements at, say, $\lambda_b = 4400$, $\lambda_r = 5500$ A. Combining (4) and (5) we have for the absorption at 4400 A.

$$A_{4400} = 5{\cdot}00\, E,$$

and for visual magnitudes

$$A_{5500} = 4{\cdot}00\, E.$$

In its modified form equation (3) now becomes

$$\left.\begin{aligned}\log r &= 0{\cdot}2\,(m_b - 5E - M_b) + 1\\ \log r &= 0{\cdot}2\,(m_r - 4E - M_r) + 1\end{aligned}\right\}, \qquad \ldots\ldots(6)$$

according as m and M are measured in photographic or visual light respectively. The distances given by (6) are still upper limits to the true values, since the star's light may have suffered, in addition, non-selective absorption not obeying (4). Spectroscopic observations can tell us nothing about this; we must turn to other observations, e.g. those of galactic clusters (Trumpler, 1930 and 1940; Zug, 1933). These agree in putting the total absorption of photographic light at about 0·7 magnitude per kiloparsec in the galactic plane, though here again the observed scatter indicates serious non-uniformity of distribution. This latter fact complicates the position so much that at this early stage few generalizations can be made about the total absorption, though it seems likely at present that the selective absorption is a considerable fraction of the whole. Nevertheless, it cannot yet be claimed that (6) establishes a reliable system of stellar distances; all that can be said is that the scale it defines is an improvement on that based on (3). The problem is by no means hopeless; it is simply much more complicated than it seemed only a few years ago, and its elucidation will tax our resources to the utmost. To the observational astrophysicist that fact should be a challenge whose acceptance will eventually provide us with the solution.

§4. CONCLUSION

At this stage of our knowledge many perplexities remain. If the various methods of attack yield apparently incompatible results, this must be attributed to two causes: that the absorbing clouds of the galaxy are as formless and as irregularly spaced as the clouds in the sky; and that our observational methods and equipment are hardly as yet capable of the star-to-star investigation which this fact necessitates. Some broad conclusions can, however, be drawn already. Firstly, our universe is only about half as big as we thought ten years ago. Secondly, it is peryaded with a cloud of gas—electrons, protons, atoms and molecules—whose presence is betrayed directly or indirectly by absorption lines in the spectra of distant stars. Finally, in its plane of symmetry there is a layer of dust a few hundred parsecs thick, the size of the dominant particles being about the wave-length of visible light. This stratum produces dimming and reddening of the light of stars in the galactic plane; and the many concentrations in it which are visible as the diffuse nebulae, bright and dark, produce heavy and irregular obscuration which immensely complicates nearly all problems of stellar statistics.

REFERENCES

§2.

ADAMS, W. S., 1940. *Harv. Announcement Card* 526, June.
BEALS, C. S., 1936. *Mon. Not. R. Astr. Soc.* **96**, 661.
DUNHAM, T., 1937. *Publ. Astr. Soc. Pacif.* **49**, 26; *Nature, Lond.*, **139**, 246.

DUNHAM, T., 1939. *Proc. Amer. Phil. Soc.* **81**, 277.
GLEISSBERG, W., 1937. *Z. Astrophys.* **13**, 255.
MCKELLAR, A., 1940. *Harvard Announcement Card* 526, June.
MERRILL, P. W., 1934. *Publ. Astr. Soc. Pacif.* **46**, 206.
MERRILL, P. W., 1937. *Phys. Rev.* **52**, 761.
PLASKETT, J. S. and PEARCE, J. A., 1933. *Publ. Dom. Astrophys. Obs.* **5**, 20.
RUSSELL, H. N.,1935. *Mon. Not. R. Astr. Soc.* **95**, 635.
STRÖMGREN, B., 1939. *Astrophys. J.* **89**, 526.
STRUVE, O., 1939. *Proc. Nat. Acad. Sci., Wash.*, **25**, 36.
STRUVE, O. and ELVEY, C. T., 1938. *Astrophys. J.* **88**, 364.
SWINGS, P., 1937. *Mon. Not. R. Astr. Soc.* **97**, 212.

§ 3.

ATKINSON, R. D'E., HUNTER, A. and MARTIN, E. G., 1940. *Mon. Not. R. Astr. Soc.* **100**, 196.
BARBIER, D. and CHALONGE, D., 1940. *C.R. Acad. Sci., Paris*, **210**, 166.
HUNTER, A. and MARTIN, E. G., 1940 a. *Mon. Not. R. Astr. Soc.* **100**, 656.
HUNTER, A. and MARTIN, E. G., 1940 b. *Mon. Not. R. Astr. Soc.* **100**, 669.
SEARES, F. H., 1940. *Publ. Astr. Soc. Pacif.* **52**, 80.
STEBBINS, J., HUFFER, C. M. and WHITFORD, A. E., 1939. *Astrophys. J.* **90**, 209.
STEBBINS, J., HUFFER, C. M. and WHITFORD, A. E., 1940. *Astrophys. J.* **92**, 193.
STEBBINS, J. and WHITFORD, A. E., 1936. *Astrophys. J.* **84**, 132.
STROHMEIER, W., 1939. *Z. Astrophys.* **17**, 83.
STRUVE, O., KEENAN, P. C. and HYNEK, J. A., 1934. *Astrophys. J.* **79**, 1.
TRUMPLER, R. J., 1930. *Bull. Lick Obs.* **14**, 154.
TRUMPLER, R. J., 1940. *Astrophys. J.* **91**, 186.
ZUG, R. S., 1933. *Bull. Lick Obs.* **16**, 119.

SOLAR PHYSICS

By A. D. THACKERAY, M.A., Ph.D.,
Solar Physics Observatory, Cambridge

§1. INTRODUCTION

The sun offers a unique opportunity of studying the physical state of a stellar atmosphere in that it is the only star which is near enough to present a directly perceptible disk. It is true that in a very few favourable cases stellar diameters can be measured by the indirect method provided by the interferometer, or, still more rarely, by photo-electric observation of the time taken by the moon to cover up the surface during an occultation (this time for the giant Antares amounts to 0^s·07). But for no star except the sun do we have any direct evidence of surface features, such as sunspots and the neighbouring heated regions known as *faculae*, of high atmospheric phenomena such as prominences, or of the existence of a surrounding corona. Much of our knowledge of the sun's outer atmosphere (chromosphere and corona) arises from the fortunate circumstance that the moon is so placed that its mean apparent diameter is a little larger than that of the sun, so that at times of total eclipse the sun's disk is neatly covered up and the much fainter outer appendages become visible, freed from the general sunlight scattered in the earth's atmosphere. Much evidence has been accumulated to prove the existence of extended chromospheres surrounding the brightest stars, and in two or three instances of eclipsing variables (e.g. Zeta Aurigae) we can study the structure of these atmospheres by watching the absorption phenomena during eclipse as a small bright component passes behind the outer atmosphere of the other (giant) component. Naturally, however, the knowledge gleaned in this way is very meagre compared with that derived from the sun at total eclipses.

The great apparent brightness of the sun compared with other stars means that for once astronomers have plenty of light available and can bring into operation the most powerful high-dispersion spectrographs available without fear of inordinate exposure times or of difficulties due to temperature shifts. It is even possible, with the help of Hale's *spectroheliograph*, to build up monochromatic images of the solar surface in light of atomic frequencies where the intensity is reduced to a small fraction of the continuous spectrum. Such photographs throw light on the variation of excitation in different regions and at different heights over the sun's surface.

Finally, the proximity of the sun gives it pride of place as the one star with which we are in touch through the medium of corpuscular as well as electromagnetic radiation. The expulsion of matter from stars, such as is observed

in extreme form in the case of novae, seems to be quite a common phenomenon with the sun as no exception, but the phenomena of aurorae and the recently discovered ionospheric disturbances provide direct evidence of unique character of particles apparently emanating from the sun and impinging on the earth's upper atmosphere. The detailed analysis of terrestrial atmospheric phenomena, and in particular the separation of effects due to particles and to ultra-violet radiation (beyond the ozone absorption limit at about 2900 A.), is likely to have a powerful influence on some astrophysical ideas about the sun which are still in a somewhat speculative stage of development.

Before describing in detail recent developments in our physical knowledge of the sun, it will help to put the problem in its true perspective to point out to what extent the sun is representative of the stars as a whole. The stars as we see them show a very great range in apparent brightness, and if we allow for their different distances to compute their *absolute magnitude* (i.e. the apparent magnitude that they would have at the standard distance of 10 parsecs, or 32·6 light-years) we find a range of at least 23 magnitudes in true brightness. Since the sun sends us about 10^{11} times as much light as a first-magnitude star, the measurement of its apparent magnitude on the stellar scale is an observation of considerable difficulty. Nevertheless, surprisingly good agreement has been attained between observations by quite different methods, and a recent compilation (Kuiper, 1938) gives a mean (photo-visual) apparent magnitude of $-26{\cdot}84$; this corresponds to an absolute magnitude of $+4{\cdot}73$. This is about four magnitudes *fainter* than the absolute magnitude of the average naked-eye star, but about eight magnitudes *brighter* than the commonest stars in space—or at least in our neighbourhood in the universe. This difference is a direct result of the great dispersion in absolute brightness. We thus see that, although the sun should be regarded in reality as an uncommonly luminous star, it is rather faint compared with those stars which can be most readily observed and with which we are consequently most familiar. A similar result holds for other characteristics; stars are known with radii as large as 500 $R_\odot$ or as small as 1/50 $R_\odot$; in mass the observed range is ?400 $m_\odot$ to 0·1 $m_\odot$; in both cases observational selection favours large values. In mean densities the observed range is from about 10^{-7} to 10^{+5}. The fact that the sun's mean density is 1·4 led at first to the belief that its interior was solid or liquid until it was realized that the high degree of ionization would permit matter to reach comparatively high densities and yet retain the essential characteristics of a perfect gas.

The total amount of energy radiated by the sun which is received per unit area and time at the earth's distance is known as the *solar constant*. By comparing this with the energy radiated by a black body at temperature T_e according to Stefan's law we can find the *effective temperature* of the sun. The most-often quoted result is that given by Abbot (1922), namely, $\sigma = 1{\cdot}940$ cal./cm^2 min., corresponding to $T_e = 5740°$ K.; a discussion (Unsöld, 1938) of recent observations leads to the revised values, $\sigma = 1{\cdot}901$, $T_e = 5713°$ K. The variations of the solar

constant, and particularly the correlation of these variations with the sunspot cycle, must be regarded with some reserve (Sterne, 1940). In this connection it should be remembered that the final value depends on a troublesome correction for absorption in the terrestrial atmosphere, which varies from day to day and from station to station; in addition, extrapolations are required in the ultra-violet and infra-red portions of the spectrum which are inaccessible to observation, amounting to 3·4 and 2·0 per cent respectively. Indirect evidence points to a variation with the solar cycle of the unobservable ultra-violet radiation; it has not yet been established with certainty that either the intensity of the solar spectrum, 2900 to 3300 A., or the terrestrial ozone absorption in this region varies regularly with the sunspot cycle.

§2. SURFACE FEATURES

Sunspots have always attracted attention as the most easily observed irregularities on the solar surface. They are of special interest as the origin of powerful magnetic fields, of the order of 2000 gauss, discovered spectroscopically by Hale as a result of his observations of vortical structure in hydrogen spectroheliograms. It is remarkable, however, that the directions of these vortices seem to be unrelated to the magnetic polarity. The vortical structure, which is supposed to be a high-level feature, follows statistically the same law as terrestrial cyclones in having constant (opposite) senses in the two hemispheres, north and south; the magnetic polarity, however, follows another law and appears to be associated with motions of charged gases at a lower level. Extensive magnetic observations of spots have been published recently by Hale and Nicholson (1938). It is a peculiar fact that the vortical motion has not been certainly detected spectroscopically through Doppler shifts; at the effective level at which absorption lines are formed the motions appear to be mainly radial (Evershed effect), the upper gases flowing inwards, the lower outwards.

Empirical laws governing the distribution and numbers of spots during the 11-year cycle (or, more strictly, the 22-year cycle, if account is taken of magnetic polarities), the behaviour during growth and decay of spots, and the similar characteristics of *faculae* are now well established. But no comprehensive theory is available to explain the facts. One conclusion however is clear; observations with spectroscope (Moore, 1932) and thermocouple (Pettit and Nicholson, 1930; Wormell, 1936) agree in assigning a temperature of about 4300° K. to the average large spot, i.e. 1400° cooler than the photosphere. The maintenance of such a cooled region for periods of time up to several months indicates that some deep-seated disturbance must be involved. A reduction of pressure in spots by a factor of about 2 is also suggested by spectroscopic observations (Moore, 1932). Thus it is natural to suppose that sunspots are cooled by the adiabatic expansion of gases rising from below. Milne (1930) has proposed a theoretical model which explains many of the observed features. Minnaert and Wanders (1932) found that the observed intensity-ratio Spot/Disk at different wave-lengths

and at different parts of the disk was consistent with radiative rather than convective equilibrium in spots, although a convective layer might exist below the surface. It now appears that in certain conditions, which may well hold in the solar atmosphere, the observed limb-darkening is consistent with the assumption that both spot and disk are in convective equilibrium (Thackeray, 1940). Until recently the general limb-darkening of the sun, which gives a direct indication of the temperature distribution with depth, had been interpreted as a proof of radiative equilibrium. The work of Unsöld (1930) on the existence of convective zones in those regions where hydrogen is half-ionized has had a powerful influence in shaping recent ideas concerning the equilibrium of the solar atmosphere (Siedentopf, 1935; Cowling, 1935).

This subject is closely associated with the problem of *granulation*, which has excited renewed interest in the last five years (Plaskett, 1936). The chief characteristics of these minute markings, which distinguish them chiefly from sunspots, are: (i) the very short lifetime of individual granules; (ii) uniform distribution over the whole solar disk; (iii) independence of the sunspot cycle.

The average size of granulation was found to be about 1 to 2 seconds of arc by observers in the last century.* While some recent observations have suggested that this result was an under-estimate, the question now seems to have been settled in favour of the earlier observers. Keenan (1938) gives a mean diameter of 1″·0 to 1″·2; we may thus picture the average granule to be about the size of the British Isles.

The contrast between granulation and background is consistent with a difference in temperature of 100 to 200° K. Recently ten Bruggencate (1939) claims to have found a granular structure in faculae with a lifetime of the order of one hour and a greater contrast than that of normal granulation.

The time of decay of individual granules was known from direct visual observations to be of the order of a few minutes, but an accurate objective measurement has been effected by a combination of very ingenious methods, involving

* The revolution in ideas about the sun that has taken place since the development of astrophysics is well illustrated by the views of two distinguished astronomers on granulation in 1865 (*Mon. Not. R. Astr. Soc.* **25**, 152):—

" I cannot divest myself of the opinion that the photosphere is really made up of clouds, and that the luminous stratum is actually constituted like our clouds, the only difference being that the clouds on the Earth are of watery drops or crystals, and in the Sun they are of some other substance in a state of suspension in the atmosphere. This opinion is enforced upon me when I observe the dissolving part of a spot, which presents all the phenomena of our cloudy sky; they are the forms of *cumuli*, of *cirri*, of *strati* . . . " (Father Secchi.)

" *Clouds*, in the ordinary sense of the word, I do not think they are. I fancy I see the key to their nature. I believe them to be *permanently solid matter*, having that sort of fibrous or filamentous structure, which fits them when juxtaposed by drifting about and jostling one against another to collect in flocks, as *flue* does in a room . . . " (Sir John Herschel.)

The point of interest is that astronomers of 75 years ago saw almost precisely as much as we do today; modern observations of granulation have not proved so superior that any startling discoveries have been made. The change that has taken place is in the background of physical ideas, determined by the interpretation of observations of other solar phenomena in the light of modern physical theory.

the production of artificial granulation with evaporating camphor powder (Grotrian *et al.*, 1936). The result is $3{\cdot}2 \pm {\cdot}05$ minutes.

§3. THE SOLAR SPECTRUM

Kirchhoff's theory of absorption and emission, which in 1859 formed the real starting point of astrophysics, is exemplified with remarkable completeness by the spectra of different portions of the sun. The continuous spectrum of the *photosphere*, or lowest layer to which we can see, is crossed by dark Fraunhofer lines produced in the upper (cooler) *reversing layer*; at the limb of the sun, where we view the still higher *chromosphere* freed from the photospheric background, emission lines appear which closely reproduce the Fraunhofer spectrum. For many years it was a puzzle why the chief difference between chromosphere and Fraunhofer spectrum lay in the strengthening of the "spark" lines—characteristic of high temperature in the laboratory—because this seemed to indicate that the top of the atmosphere was hotter than the bottom. Saha's ionization theory (1920) showed that the lower pressure of the higher layers could more than counterbalance a lower temperature in strengthening the lines due to ionized atoms. Ionization theory, which thus largely originated from a strictly solar problem, has played a fundamental rôle ever since in the analysis of celestial spectra.

The Fraunhofer spectrum has been observed and catalogued in great detail by observers at the Mount Wilson Observatory in the *Revised Rowland Table of Solar Wavelengths* which contains over 20,000 lines between 2975 and 10,000 A. (St. John *et al.*, 1928). The ultra-violet limit is set by terrestrial ozone absorption, the infra-red by the sensitivity of the photographic plate. An extension to about 13,000 A. has been achieved by H. D. Babcock, but the results are not yet fully published. While only 55 per cent of the lines have been identified, the remainder are all very weak (and probably mainly due to molecules), and there is no reason to suppose that any of them are due to any undiscovered terrestrial element. The position with regard to the identification of elements in the sun is subject to revision from time to time as new solar and laboratory data come to hand; 63 elements are now known to be definitely present; elements that have recently been included in this number are tin, osmium and iridium (Moore, 1937 and 1939; Abetti, 1939).

The problem of the quantitative abundance of elements is a much more difficult one than the mere proof of their presence. For a complete solution full knowledge of atomic energy levels and transition probabilities is required Nevertheless great progress has been achieved, and separate analyses, which give roughly concordant results, have been given for the Fraunhofer spectrum (Russell, 1929), for sunspots (Moore, 1932) and for the chromosphere (Menzel, 1931). It is found that with a few notable exceptions the relative abundance of elements is very similar to that observed in the earth's crust and in meteorites. In particular, the elements of even atomic number are on the average ten times as abundant as the odd. Lithium, found to be much rarer in the sun than in

the earth, is an interesting case; according to current theories of energy generation in stellar interiors due to nuclear transformation, this element would be used up in a chain reaction in a very short space of time. But the most important outcome of this work has been the recognition of the enormous preponderance of hydrogen. While the quantitative abundance is still uncertain, it appears that in the solar atmosphere hydrogen atoms outnumber all the metals put together in the approximate ratio of 1000:1. The theory of stellar interiors yields two alternative solutions for the solar hydrogen abundance of 33 and 99·5 per cent (by mass) (Eddington, 1932). The fact that hydrogen is not nearly so abundant on a small planet like the earth need occasion no surprise, as the greater part was probably lost in the early stages of terrestrial existence. It may be mentioned in passing that supporting evidence is found in the abundance of hydrogen compounds (NH_3, CH_4) discovered recently in the spectra of atmospheres of the heavy planets Jupiter, Saturn, Uranus and Neptune (Russell, 1936).

The above investigations have depended on visual estimates of intensity, and there is room for considerable improvement in the use of "total absorptions" determined photometrically (Allen, 1935); such measures should theoretically be independent of resolving power, provided there is reasonable freedom from blending with neighbouring lines. But the greatest astrophysical need in the analysis of line-intensities is for empirical or theoretical transition-probabilities (*f*-values). A start has been made with neutral lines of Fe and Ti, observed in absorption with an electric furnace providing the excitation (R. B. and A. S. King, 1938). The solar intensities of these lines suggest that the atoms follow a Boltzmann distribution corresponding to a temperature of 4400° K. Similar results have been obtained from analyses of molecular band intensities (Birge, 1922; Adam, 1938). The fact that this *excitation* temperature is so far below the effective temperature has not yet been fully explained, although a qualitative explanation is readily found in the lower temperature of the reversing layer compared with the photosphere. Theoretical *f*-values within "supermultiplets" have been calculated by Goldberg (1935, 1936), which yield similar results for the solar excitation temperature (Menzel *et al.*, 1938). The most recent determination of the excitation temperature from a CN band yields the result 4490 ± 100°K. (Blitzer, 1940).

The *colour temperature* of the sun, i.e. the temperature of a black body that describes most closely the observed distribution of energy in the continuous spectrum, has been found to be about 7100° for the visual region (Mulders, 1936); this high value has been reached as a result of applying corrections for absorption lines which tend to crowd together in the violet; the colour temperature in the ultra-violet seems to be as low as 4800°, but until the theory of the continuous absorption coefficient in stellar atmospheres has been clarified, the physical meaning of these colour temperatures must remain somewhat obscure.

The theoretical interpretation of line-intensities has been aided by the introduction of the conception of the "curve of growth", i.e. the functional relationship

between the equivalent width W and the number of effective absorbing atoms Nf (van der Held, 1931; Minnaert, 1934). Faint lines, whose width is determined by Doppler broadening, obey the law $W \propto Nf$, while strong lines, dominated by "natural" radiation damping, obey the law $W \propto \sqrt{Nf}$. Observations of the curve of growth enable one to derive an empirical astrophysical mean value of the atomic damping constant. This, for sun and stars alike, turns out to be about ten times the classical value (Allen, 1935). Since, on the quantum-mechanical model, the damping constant is a function of the lifetimes of the excited atomic states, the discrepancy is not altogether surprising. A *kinetic temperature*, which describes the extent of Doppler broadening, may be derived from the observed curve of growth, and Rubinstein (1940) has in this way found the value $5400 \pm 1300°$ K.

The resolving power of solar spectrographs is sufficient to enable the shapes or contours of the broader absorption lines to be observed, and this problem has received considerable attention from both theoretical and observational points of view (Redman, 1939). In particular, the variation of lines across the disk from centre to limb offers some hopes of analysing the physical state of the sun's atmosphere at different depths (Plaskett, 1931; Minnaert and Houtgast, 1936). The situation here is a complicated one, but one difficulty has been largely solved; theory predicts that the centres of strong absorption lines should be nearly perfectly black, but early observations scarcely ever gave central intensities less than 20 per cent of the adjacent continuous spectrum. By careful spectrophotometric technique and elimination of effects due to scattered light and instrumental resolving power, the centres of strong metallic lines have been shown to be of the order of 1 to 2 per cent (Redman, 1935, 1937). The Balmer lines of hydrogen (and to a less extent the H and K lines of Ca II) form a notable exception in having central intensities of 15 per cent (Thackeray, 1935); this may be due to the existence of the strong Lyman lines in the far ultra-violet and to the mechanism of fluorescence (Woolley, 1934).

§4. WAVE-LENGTHS OF SOLAR ABSORPTION LINES

The wave-length and strength of a line together give the main clues to its identification with an element. Any small residual shift compared with a terrestrial source is commonly attributed to a Doppler effect due to motion in the line of sight. A clear verification of this principle is afforded by a comparison of the sun's eastern and western limbs, which show shifts consistent with a rotational velocity (of about 2 km./sec.), in agreement with that suggested by direct observation of the daily motion of spots and faculae across the sun's disk. Spots are not seen above 40° latitude, and spectroscopic observations can be carried out much nearer to the poles. An important result is that the sun does not rotate as a rigid body; the equator rotates faster than the regions of high latitude. The angular velocity of the higher atmospheric layers also appears to be more rapid than that of spots. These results belong to a branch of astrophysics which might be called

Solar Meteorology, the theory of which is still in its infancy. Little progress has been achieved since a pioneer attempt to combine observations of solar rotation and of sunspot vortices in a single hydrodynamical theory (Bjerknes, 1926). Observations of solar rotation have been collected in a recent survey by De Lury (1939).

The theory of relativity predicts a general shift to the red of all lines in the solar spectrum, of an amount corresponding to a velocity of 0·63 km./sec.; this is due to the loss of energy of quanta in escaping from the sun's gravitational field. This problem has been the subject of extensive investigations by St. John (1928) and Evershed (1936, 1938). St. John found residual shifts between strong and weak lines which could best be interpreted on the hypothesis of convection currents, the lower levels rising and the upper falling. Observations at the centre of the disk are therefore less suitable for studying the relativity shift than those at the limb, where convection currents will lie across the line of sight. Evershed's observations indicate a limb-shift considerably in excess of Einstein's value, and this problem has not yet been solved. It is significant that Evershed's results have been substantially confirmed by measures during eclipse, when effects due to scattered light are practically eliminated (Royds, 1937). The variation of the shift across the disk might (on St. John's hypothesis) be expected to follow a simple $\cos\theta$ law; but it has been shown that to explain the observed variation it is necessary to take into account the higher effective level of the limb-lines and the possible variation with height of the velocity of the convection current (McCrea and Mitra, 1934). This explanation does not, however, account for the residual limb-shift. An unexplained shift of the absorption lines relative to the chromospheric emission lines observed at the 1936 total eclipse may throw some light on this difficulty (Thackeray, 1937).

At this stage two further difficulties provided by observations of the extreme limb may be mentioned: (*a*) the limb-darkening near the extreme limb appears to be independent of wave-length (Wesselink, 1940); (*b*) faint lines are strengthened at the limb in a manner that is difficult to explain by current theories (Adam, 1937 and 1940). The problem of contours of lines at the extreme limb has been considered theoretically by Woolley (1937).

§ 5. THE EQUILIBRIUM OF THE CHROMOSPHERE

Gravity at the sun's surface amounts to 28*g*, and a solar atmosphere in hydrostatic equilibrium at 6000° K. acted upon solely by this force would extend to a height of only about 10 km.; in such circumstances the chromosphere would be below the limit of observation. The strongest emission lines, due to ionized calcium and hydrogen, are observed extending to a height of 10,000 km. or more above the limb. Radiation pressure is the most natural force to invoke as an explanation of this distension of the atmosphere, and Milne in a series of classical papers has worked out the equilibrium of a calcium chromosphere, partially or completely supported by radiation pressure. In such equilibrium the central

intensity within the calcium lines should bear a simple relation to the excited lifetime of the atom, and the value obtained astrophysically agrees well with the laboratory value of 10^{-8} sec. (Milne, 1924). An interesting by-product of the investigation was to show that such an atmosphere would be stable for downward chance velocities, unstable for upward velocities; this is a consequence of Doppler effect causing the atoms to lose the shelter of the strong absorption core and to become exposed to strong radiation pressure in the wings (Milne, 1926). The limiting velocity of escape was calculated to be about 1800 km./sec., and this agrees well with the observed time-lag between solar and terrestrial phenomena which are attributed to corpuscular streams.

Chandrasekhar (1933, 1934) modified Milne's theory by taking into account the patchy nature of the solar surface; a sinusoidal variation of the intensity, with a "wave-length" of 20,000 km. and an amplitude of ± 33 per cent, must be regarded as a distinct improvement on the assumption of a uniform flux of radiation at all points. Of the resulting trajectories some are periodic and others aperiodic; the latter are compared with the eruptive "spikes" in the chromosphere, and the theoretical velocity of 30 km./sec. agrees with observation. The envelopes of these trajectories define the chromospheric density distribution, but here the agreement with observation is less satisfactory.

The great difficulty that faces all theories which ascribe the support of the chromosphere to radiation pressure is that there is no evidence that calcium really rises to greater heights than other elements; it is a favoured element in that its resonance lines, which are particularly strong, are directly accessible to observation. But they also lie in a part of the solar spectrum where the intensity of radiation is nearly a maximum. Many other elements, such as Fe^+ and particularly hydrogen and helium, have their resonance lines in the far ultra-violet, and if the sun's energy corresponds even remotely to that of a black body at 6000° K., radiation pressure would be quite insufficient to support these elements. It might be thought that the few favoured elements, Ca^+, Sr^+, Ba^+, subjected to intense radiation pressure, might communicate their momentum to others through the action of viscosity. It has, in fact, been shown that the Ca^+ atoms in prominence could "drag" hydrogen and other elements with them (McCrea, 1935), but these calculations referred mainly to motions of isolated clouds of prominence matter, and it seems impossible that selective radiation pressure could account for the complete support of the chromosphere.

In order to cope with the difficulty of the equilibrium of the hydrogen chromosphere, McCrea (1929) introduced the conception of turbulence. In this theory the atmosphere is considered to consist of volume-elements within which the atomic motions are characterized by a temperature parameter T^* of the order of 6000°, but the velocities of the individual volume-elements (assumed to follow a Maxwellian distribution) by a much higher temperature parameter T. The emission cores of the Ca^+ lines have a width which corresponds to a mean turbulent velocity of 18 km./sec.; this leads to $T^* = 18{,}600°$, and such a temperature is suffi-

ciently high to account for the observed chromospheric density gradients. No specific explanation of the origin of such turbulent velocities can be offered, but they may well arise in some deep-seated layer (the convective zone?), and it may be pointed out that the assumption is fully consistent with the observed facts that the sun's surface is in a state of continual agitation and that prominences with velocities of several hundred km./sec. are not uncommon. The procedure, frequently adopted by theorists for the sake of mathematical convenience, of assuming the solar atmosphere to be stratified in plane parallel layers of constant temperature and density has doubtless been adopted too readily, and in recent years there have been many signs of a healthy reaction.

There have been considerable improvements in observational technique in the matter of chromospheric gradients. Thanks to the introduction of photographic photometry, the first quantitative measures became available (Davidson *et al.*, 1928; Pannekoek and Minnaert, 1928). At the beginning or end of the total phase of a solar eclipse the chromospheric arc is so narrow that a slit can be dispensed with; the emission lines then appear as arcs whose lengths are closely associated with the height above the photosphere to which they can be observed. Irregularities in the moon's surface introduce some complications, but the method is essentially very simple, and has been used extensively, especially by Mitchell (1930). In the moving-plate method, due to W. W. Campbell, only a small portion of the crescent is photographed, and the plate is moved uniformly in such a way that the spectrum is broadened (at right angles to the dispersion) into a series of lines whose intensity gradients correspond to density gradients in the solar atmosphere; this simple correspondence is a consequence of the small extent of the chromosphere compared with the solar radius. An extensive analysis of Campbell's photographs (which were not calibrated for photometry) has been performed by Menzel (1931), who finds indications that turbulence increases with height. In this connection it is relevant to quote the relatively small value of 1·5 km./sec. found for the turbulent velocity of the reversing layer from cores of absorption lines (Allen, 1937). A recent development, due to Menzel, is the jumping-film method—a compromise between the two methods described above; here the complete arcs are photographed in rapid succession on cinematograph film. Results obtained by this method at the 1932 eclipse (through thin cloud) have been discussed by Menzel and Cillié (1935), who give gradients for individual elements. The extensive material obtained by the 1936 Harvard eclipse-expedition has not yet been reduced sufficiently for publication.

To summarize the subject of chromospheric gradients, it may be said that: (1) the density falls off much more sharply for neutral than for ionized elements, in accordance with ionization theory; (2) there is no indication of a sorting out of elements according to atomic weight or selective radiation pressure; (3) in many cases the gradient cannot be represented by a single exponential factor: the gradient usually *diminishes* in the higher layers; in the case of helium there is some evidence for *stratification* at a high level.

The problem of helium in the chromosphere is a peculiar one. The only lines of this element in the accessible region have excitation potentials of about 20 electron-volts, much higher than those of any other lines in the solar spectrum—with one notable exception, due to ionized helium; the line λ4686 (excitation potential 48 volts) is also observed faintly in the chromosphere. It is impossible to account for these lines on any assumption involving thermodynamical equilibrium. The explanation that is most favoured is that they arise from electron capture by singly or doubly ionized helium atoms, the source of the high-energy excitation being supplied by an excess of ultra-violet radiation in the region 200 to 500 A. which is somehow permitted to escape from the photosphere. There are indications that the regions of high excitation ("hot spots") are strongly localized.

§ 6. THE MOTIONS OF PROMINENCES

Two developments in observational technique have combined to stimulate interest in the study of the motions and forms of prominences. Ever since 1868 it has been realized that prominences, whose radiation consists mainly of a few strong-line frequencies (notably Hα), can be readily observed outside eclipse with the spectroscope, which frees the monochromatic light from the general background of scattered sunlight. But it is essential to use a slit-spectroscope, and consequently only a small section of a prominence can be observed at one time. The spectroheliograph enables one to build up a complete photograph, but this process tends to be slow with the powerful instruments that are normally used. In 1924 Hale perfected his spectrohelioscope wherein an oscillatory motion of the slit permits visual examination of a considerable area at a time of prominences and of surface markings on the disk. A particularly ingenious arrangement of rotating rectangular prisms eliminates the necessity of an oscillating slit (which tends to introduce mechanical difficulties) (Hale, 1929). In this form the instrument has been universally adopted for a programme organized by the International Astronomical Union, whereby the sun is kept under practically continuous observation, in Hα light, at stations distributed as uniformly as possible in longitude (d'Azambuja, 1939).

A second development has been the introduction of motion-picture technique. The spectrohelio-kinematograph of the McMath-Hulbert Observatory is capable of taking Hα or Ca^+ pictures in rapid succession (down to intervals as small as 10 seconds) which can either be measured at leisure or combined into a moving-picture in which the increase in the time-factor is normally about 500 (McMath, 1937).

Another method has been used by Lyot (1939). The coronagraph that he has constructed with such care and precautions against scattered light, used at a height of over 9000 ft. on the Pic du Midi with the help of suitable red filters, is capable of directly recording prominences *without a spectroscope*. The method has the advantage of attaining the optimum definition permitted by atmospheric "seeing". In the spectrohelio-kinematograph a rather wide slit is used, partly for reasons of

speed, but chiefly because otherwise Doppler motions of rapid prominences might cause them to be completely unrecorded; consequently the finest detail tends to be slightly blurred.

Such films have had an inspiring effect on all astrophysicists who have had the good fortune to see them, and have even stimulated hydrodynamical experiments in search of analogous motions in the laboratory (Zanstra, 1939). One new observational feature has been emphasized which would have been very difficult to determine by purely visual means; motions in prominences are observed in all directions, but predominantly *downwards* towards the sun's surface; some estimates put the proportion of matter so moving as high as 90 per cent. Early visual observations divided prominences into two broad types, quiescent and eruptive. The brilliance and spectacular nature of the latter type easily attracts attention, and here the predominant motion is upwards, suggesting the repulsive force of radiation pressure. It has been known for some time that sunspots often act as "centres of attraction" in sucking down prominence matter (Keenan, 1935), but the frequency of downward-moving faint streamers is a new feature revealed by these films. The accelerations are found to vary markedly from zero to values several times that due to gravity (McMath, 1937; Sawyer, 1939). Pettit finds that eruptive prominences move with uniform velocities upwards, but are subject to abrupt changes (1932); his attempts to find a simple relation between the velocities before and after the change do not seem to be justified when the data are examined statistically (Hulme, 1939). The general rate of increase seems to be faster than that predicted by Milne's theory of the ejection of Ca ions. With the accumulation of observations a more detailed subdivision into types has been suggested; besides eruptive and quiescent, forms are classified as active, spot and tornado. The active type is the commonest, the tornado the rarest, comprising 72 and 1 per cent respectively. In "surges," a short-lived brightening develops and contracts again along the same path. The *coronal streamers* are perhaps the most puzzling phenomena of all; these appear to originate in some invisible source, occasionally higher than 100,000 km., and to pour out luminous streams in various directions towards the surface.

This brief summary of the rich material contained in recent observations is sufficient to indicate the complexity of the phenomena. No real understanding of the laws governing the motions of prominences has been attained, and one can only conjecture that they represent the consequences of an interplay between forces of an electromagnetic character, of radiation pressure and of gravitation, in which the first two forces are subject to rapid local variations.

A criticism that is sometimes brought against the conception of electric forces playing an important rôle is that ionized calcium and neutral hydrogen show identical motions as far as observations can probe. This difficulty may have been overstressed; as mentioned previously, viscosity has been shown to be capable of causing one element to drag another with it; and secondly, the observed emission might be attributed entirely to a stream of *electrons* in the process of

capture by doubly ionized calcium and by hydrogen nuclei. A point that has to be continually borne in mind is that the observations may equally well be interpreted by changes in the state of excitation as by motion of matter. It is possible that in course of time observations of Doppler effect at the limb and over the disk will provide a decisive answer to this question.

§7. CHROMOSPHERIC ERUPTIONS

The occasional intense brightenings that are recorded on Hα or Ca^+ spectroheliograms have been known for many years. Since they last for comparatively short periods, varying between a few minutes and an hour or two, it is no easy matter to record in detail the rapid changes that take place, especially when observations are apt to be interrupted by clouds. The unusually high maximum of solar activity in 1937–8, coupled with an international programme of spectrohelioscope observations, has produced a rich harvest of observational material; this was urgently needed in view of the recent discovery of the association between solar and ionospheric disturbances (Mögel, 1930; Dellinger, 1935; Richardson, 1935 and 1937; Newton and Barton, 1937; Waldmeier, 1938).

This manifestation of a solar-terrestrial relationship may be summarized as follows:—

(i) The E and F_1 layers of the ionosphere undergo a regular variation with the sunspot cycle, which indicates a total change of about 150 per cent in the rate of electron production (Appleton, 1939). This may be attributed to a corresponding change in the intensity of a part of the sun's ultra-violet spectrum.

(ii) There is an even greater variation in the ionization in the anomalous F_2 layer during the sunspot cycle.

(iii) Temporary ionospheric disturbances, which exhibit a great range in degree of violence, can usually be traced directly or indirectly to chromospheric eruptions. (*a*) Ionospheric "irruptions" or, in other words, the well-known radio fade-outs, closely coincide in time with solar eruptions. In particular, fade-outs are confined to the earth's sunlit hemisphere, and their frequency of occurrence increases with the solar altitude (Giovanelli and Higgs, 1939). (*b*) Ionospheric storms occur during and after terrestrial magnetic storms, especially in the F_2 layer. The association of aurorae and magnetic storms with the existence of large sunspots near the centre of the sun's disk has been known for many years. It is now known that solar eruptions are statistically related to magnetic storms (Newton, 1939). In fact, the correlation with eruptions seems to be closer than that with sunspots.

The search for time-coincidences in solar and terrestrial phenomena is complicated by the fact that ionospheric disturbances, at least in the F_2 layer, can be caused by corpuscular streams as well as by electromagnetic radiation. The former take 1 to 1½ days to reach the earth, the latter 8 minutes. Thus the need for an infallible method of distinguishing between effects due to particles and to radiation is imperative. The association with sunspots must be regarded as an

indirect one, for it has become increasingly clear that the primary cause of ionospheric disturbances lies in an outburst of unobservable ultra-violet radiation, accompanied by a stream of ions and electrons ejected more or less radially from an active area of the sun's surface, which may perhaps include a large spot. It has, however, been shown that the statistical probability of the occurrence of an eruption near a spot is proportional to the spot-area, but independent of the strength of the associated magnetic field (Giovanelli, 1940). The hydrogen brightenings that can be observed must also be regarded as secondary effects, since the increase in Hα light amounts to only a fraction of one per cent of that from the whole solar surface (Dobbie *et al.*, 1938), while the increase in ionizing radiation must be several hundred per cent.

In view of the extremely ephemeral character of eruptions, their spectra are of special interest. Photographs have been obtained which suggest clues to the nature of the unobservable ultra-violet spectrum (Richardson and Minkowski, 1939; Allen, 1940).

The spectra of different eruptions seem to be essentially the same, and very similar to that of the chromosphere in consisting of bright lines of H, He and Ca^+. Evidence has also been found for the brightening of Fe^+ and Ti^+ lines, especially the former. We may therefore expect strong emission in line-frequencies at 1215 A. (Lyman α) and at 584 A. (He). While these lines may contribute to the variation of ionization in the earth's atmosphere, and it has even been thought possible that the various layers in the ionosphere are due to the penetration of an ultra-violet line-spectrum, it is doubtful whether λ584 could penetrate to the low depth where the effects responsible for fade-outs mainly occur. A recent advance, which may help in the analysis of the ultra-violet solar spectrum, has been the successful taking of spectroheliograms in the light of the infra-red He line at 10,830 A. (d'Azambuja, 1939). This line has the unusually high excitation potential of 20 volts; its presence in absorption in regions near sunspots must indicate radiation of very high frequency, or perhaps high-speed corpuscular streams. These infra-red spectroheliograms may in a sense be regarded as *negative* pictures of the sun's ultra-violet radiation. There is some evidence that the He absorption increased in the period 1934–8 from minimum to maximum solar activity.

The question of what part a continuous ultra-violet spectrum may play in these effects is still uncertain, but any continuous spectrum in the visual region accompanying eruptions must be very weak. There is, however, the possibility that at certain wave-lengths variations in the continuous absorption coefficient may lead to ultra-violet "leaks" from active regions of the solar surface. The further analysis of radio-observations, of the night-sky and dawn-flash spectra and of comets' tails probably offer greater hope of understanding the ultra-violet spectrum of chromospheric eruptions than direct observation of the sun. It is an unfortunate circumstance that eruptions take place at such a level that they cannot be isolated, like a prominence, from the photospheric background. Eruptions

can be seen close to the limb, and in such cases they are frequently observed to eject eruptive prominences; this latter behaviour is not so well defined that if an eruptive prominence is observed without a bright "root" it can be inferred with certainty that there is such a root behind the limb. It has been claimed that in order to be effective in producing a fade-out, the root of the eruption must be visible.

A fact which must be taken into account by any theory of eruptions is that the bright area on the sun's disk shows no appreciable radial velocity; motion in the line of sight is usually confined to the dark absorption markings which can be identified with prominences; 65 per cent of these dark markings show downward velocities (Newton, 1934), in agreement with the preponderance of apparent downward motions observed at the limb.

With reference to the connection of solar activity with magnetic storms, the persistence of the 27-day recurrence of terrestrial storms throughout sunspot minimum remains a feature of peculiar interest. The evidence for believing that the 27-day period has a real connection with the rotation of disturbed solar areas has been strengthened by the discovery of a small change in the period with the solar cycle which is consistent with the change in latitude of spots and consequent change in rotation period (Archenhold, 1939).

§ 8. THE CORONA

The total amount of light emitted by the corona amounts to about half that of the full moon, or one millionth of that emitted by the sun. While the brightness shows fairly definite indications of varying from eclipse to eclipse, such variations have not been found to be correlated with the sunspot cycle (see Dyson and Woolley, 1937). The shape of the corona, however, undergoes secular changes; at maximum the form is roughly symmetrical, while at minimum, long equatorial streamers stand out in contrast to short "plumes" near the poles. This law is subject to some irregularities, e.g. at one eclipse opposite sides of the sun showed opposite coronal types.

If it were not for the fortunate circumstance of total eclipses we would probably have no knowledge of the existence of the corona. Many abortive attempts have been made in the past sixty years to observe it without an eclipse. Shortly after a "proof" had been given of the impossibility of this task owing to scattered sunlight, the problem was brilliantly solved by Lyot (1931), who first detected lines in the coronal spectrum on 1930 August 8. Lyot's success, which will stand as one of the finest scientific achievements of this century, is to be ascribed mainly to very careful technique in construction and design which reduced to a minimum light scattered or diffracted from scratches on glass surfaces, from edges of lenses and from the walls of the coronagraph. An interesting alternative method, which employs television scanning and high amplification has been applied by Skellett (1940). The high contrast thus obtained enables prominences and the brightest coronal features to be detected. If the definition of the image can be improved, this method may eventually supplant that of direct photography.

Lyot's work has led to the discovery of a number of new coronal emission lines, chiefly in the infra-red; in particular, he has found a pair of lines at 10,747, 10,798 A. which are even stronger than the well-known λ5303 (Lyot, 1939). In the violet and ultra-violet, difficulties due to scattered light still give the advantage to eclipse observers. Dunham (1939) found a new line at λ4412 at the 1937 eclipse; the fact that it then appeared stronger than other well-known lines emphasizes the variability of the coronal spectrum.

The distribution of coronal light in its various line-frequencies exhibits some well-defined differences; such data give valuable clues to the origin of these unidentified lines, and Lyot's day-to-day observations are of special importance in this respect. For instance, it appears that λ5303 has a different distribution both from the red line λ6374 and from the strong infra-red pair. Hopes that they might be identified with some forbidden transitions have been disappointed in spite of a very thorough theoretical investigation and of the observational discovery of the auroral line 5577 [O I] in the high chromosphere by H. D. and H. W. Babcock (1934). The lines have a breadth of about 1 A. or even more, indicating a very short lifetime, while forbidden lines, originating from metastable states with lifetimes of the order of one second, are extremely sharp. It has therefore been suggested that they might be due to doubly excited atoms, e.g. helium, or to negative ions (Goudsmit and Wu, 1934). Theoretical work has indicated that negative ions should possess only one stable state and should, therefore, give rise to only a continuous spectrum. Nevertheless it has been calculated that H ions are relatively abundant in the high chromosphere (Wildt, 1939), and the possible existence of unstable atomic levels which would give rise to broad lines should not be forgotten. Experimental and theoretical work on this point are urgently required. The fact that the coronal lines, almost the only spectrum in astrophysics that remains unidentified, have only once been detected in any other celestial object—in the abnormal Nova RS Ophiuchi (Adams and Joy, 1934)—is a sufficient indication of the peculiar interest of this problem.

The distribution of λ5303 radiation has been the subject of a special investigation by Waldmeier (1939). He finds the axis of symmetry to coincide with the pole of rotation. The axis of the sun's general magnetic field, which has been believed to be responsible for the equatorial streamers, was found at the time of discovery to deviate from the rotation axis by about 6°. But the need for new observations of the general magnetic field has become urgent in view of the fact that a remeasurement of the old plates has shown that its strength can hardly be as large as formerly supposed (Evershed, 1939). Waldmeier has also announced that in certain sunspot regions λ5303 can become very intense and attain a breadth of 3 A.; in such regions the line shows some self-absorption. This latter observation should finally dispose of the belief that the line is a forbidden transition.

The coronal emission lines are observed extending to a height of about 5 minutes of arc from the sun's limb. Within this distance there is also a pure

continuous spectrum without absorption lines. The rate of falling-off in intensity (as a function of the distance r from the sun's centre) varies from r^{-7} in the inner corona to r^{-2} in the outer corona. Outside 5 minutes, a normal Fraunhofer absorption spectrum appears in which the lines are shallower, but of about the same width as in the solar spectrum (Grotrian, 1934). In addition, the continuous spectrum is radially polarized; the proportion of polarization reaches a maximum of about 30 per cent at a distance of 10 minutes (Young, 1911; Johnson, 1934). The colour of the corona is essentially the same as that of the sun.

The combination of all these observational facts into a single theory constitutes one of the major unsolved problems in solar physics. The presence of polarization suggests a scattering atmosphere, but scattering by small particles such as ions or electrons seems to be ruled out by two considerations: (i) the absence of a Rayleigh "blue"; (ii) the presence of Fraunhofer absorption lines in the outer corona; the thermal motions of electrons would be so great that they would broaden the lines to the point of invisibility. On the other hand, scattering by large particles raises the question of how such particles could exist or arrive so near the sun without volatilizing.

Minnaert (1930) suggested that the corona emitted a reddish spectrum of its own (unpolarized) which counterbalanced a blue (polarized) spectrum due to reflection by atoms; in addition, he postulated a white, polarized spectrum, without absorption lines due to reflection by electrons. Subsequent observations that the proportion of polarized light is roughly independent of λ has weakened support for this theory (Dufay and Grouiller, 1933).

An alternative hypothesis has been proposed by Grotrian (1934). According to this the Fraunhofer spectrum of the corona is due to scattering by dust particles of dimensions 3μ or greater, and there is a superposed continuous spectrum due to electron-scattering. No very convincing answer, however, is offered to the problem of the volatilization of the cloud. But the phenomena of the zodiacal light and *Gegenschein* provide direct evidence of the existence of such a cloud in the neighbourhood of the earth, and observations of the zodiacal light during total eclipse, such as have been attempted by Honda (Stratton, 1937), are of considerable interest in this respect, in that they suggest that such a cloud permeates the whole system. The difficulty with regard to volatilization is partially solved if it is assumed that the dust particles are falling into the sun, and accretion of interstellar matter has even been brought forward as a significant factor in stellar energy generation (Lyttleton and Hoyle, 1939). It has been shown that the dynamical effects of radiation are such that the sun would draw inwards such a cloud of dust, and that the whole solar system would be swept clear in a period equal to the age of the earth (Robertson, 1937). It is therefore a matter of some perplexity that the Fraunhofer lines of the outer corona have been found to be shifted to the red by about 20 km./sec., which can be most readily interpreted as a velocity of *expansion* (Moore, 1933). Accepting this expansion, we are faced by the unnatural supposition that the sun is ejecting solid dust particles, or

alternatively that the corona is composed of ions and electrons, but that it is deficient in Rayleigh blue for some such reason as Minnaert has suggested. One cannot conclude these remarks on the spectral features of the corona without registering some surprise that any emission lines *can* appear in a region where none of the familiar Balmer lines or Ca^+ (H and K) are detected.

As in the problem of prominence-motions, it is clear that the corona must be regarded as a complex dynamical problem; the rôle played by electromagnetic forces is emphasized by the shape of the coronal streamers and the formation of arches round prominences and spots. Electrodynamic theories have been discussed by Rosseland (1933) and Kiepenheuer (1935). The latter has suggested that the change in form of the corona from sunspot maximum to minimum is due to a variation in velocity of expulsion of ions and electrons from the sun's surface.

§9. SUMMARY

To review briefly the salient points in recent progress in Solar Physics, the most striking general feature has been the transition from the theoretical concept of a uniform static solar surface to that of a rapidly changing one. The calm sea has become a stormy ocean. While many astronomers, particularly observers, have never lost sight of the disturbed character of the sun, the period 1920–30, which marked the development of radiative equilibrium at stellar boundaries and the beginning of the theory of line-contours, demanded the simplifying approximation of a stratified surface. The turn of the tide is exemplified by the renewed interest in the problem of granulation and the associated question of convection, by the appeal to turbulence for an explanation of chromospheric heights and other effects, and finally, by the recent observations of chromospheric eruptions, motions of prominences and of the corona. In Solar Physics, the period 1930–40 may well be said to have been a dynamic decade.

REFERENCES

§1

Abbot, C. G., 1922. *Smithsonian Ann.* **4**.
Kuiper, G. P., 1938. *Astrophys. J.* **88**, 429.
Sterne, T. E., 1940. *Observatory*, **63**, 183.
Unsöld, A., 1938. *Sternatmosphären* (Berlin), p. 31.

§2

ten Bruggencate, P. and Grotrian, W., 1936. *Z. Astrophys.* **12**, 323.
ten Bruggencate, P., 1939. *Z. Astrophys.* **19**, 59.
Cowling, T. G., 1935. *Mon. Not. R. Astr. Soc.* **96**, 15.
Hale, G. E. and Nicholson, S. B., 1938. *Publ. Carneg. Instn.* no. 498.
Keenan, P. C., 1938. *Astrophys. J.* **88**, 360.
Milne, E. A., 1930. *Mon. Not. R. Astr. Soc.* **90**, 487.
Minnaert, M. and Wanders, A. J. M., 1932. *Z. Astrophys.* **5**, 297.
Moore, C. E., 1932. *Astrophys. J.* **75**, 222, 298.
Pettit, E. and Nicholson, S. B., 1930. *Astrophys. J.* **71**, 153.
Plaskett, H. H., 1936. *Mon. Not. R. Astr. Soc.* **96**, 402.
Siedentopf, H., 1935. *Astr. Nachr.* **255**, 157.
Thackeray, A. D., 1940. *Observatory*, **63**, 175; *Mon. Not. R. Astr. Soc.* **100**, 614.

UNSÖLD, A., 1930. *Z. Astrophys.* **1**, 138; **2**, 209.
WORMELL, T. W., 1936. *Mon. Not. R. Astr. Soc.* **96**, 736.

§ 3

ABETTI, G., 1939. *Trans. Int. Astr. Un.* **6**, 64.
ALLEN, C. W., 1934. *Mem. Commonw. Solar Obs. Aust.* no. 5.
BIRGE, R. T., 1922. *Astrophys. J.* **55**, 273.
BLITZER, L., 1940. *Astrophys. J.* **91**, 421.
EDDINGTON, A. S., 1932. *Mon. Not. R. Astr. Soc.* **92**, 471.
HELD, v. D., 1931. *Z. Phys.* **70**, 508.
KING, R. B. and KING, A. S., 1938. *Astrophys. J.* **87**, 24.
MENZEL, D. H., 1931. *Publ. Lick Obs.* **17**.
MINNAERT, M., 1934. *Observatory*, **57**, 328.
MOORE, C. E., 1932. *Astrophys. J.* **75**, 222, 298.
MOORE, C. E., 1937. *Astrophys. J.* **85**, 79.
MOORE, C. E. , 1939. *Proc. Amer. phil. Soc.* **81**, 125.
PLASKETT, H. H., 1931. *Mon. Not. R. Astr. Soc.* **91**, 870.
REDMAN, R. O., 1935. *Mon. Not. R. Astr. Soc.* **95**, 742.
REDMAN, R. O., 1937. *Mon. Not. R. Astr. Soc.* **97**, 552.
REDMAN, R. O., 1938. *Mon. Not. R. Astr. R. Soc.* **98**, 311.
RUBINSTEIN, P. J., 1940. *Astrophys. J.* **92**, 114.
RUSSELL, H. N., 1929. *Astrophys. J.* **70**, 11.
RUSSELL, H. N., 1935. *The Solar System and its Origin.*
SAHA, M. N., 1920. *Phil. Mag.* **40**, 472, 809.
ST. JOHN, C. E., MOORE, C. E., WARE, L., ADAMS, W. S. and BABCOCK, H. D., 1928. *Publ. Carneg. Instn.* no. 396.
THACKERAY, A. D., 1935. *Mon. Not. R. Astr. Soc.* **95**, 293.
WOOLLEY, R. v. D. R., 1934. *Mon. Not. R. Astr. Soc.* **94**, 631.

§ 4

ADAM, M. G., 1937. *Mon. Not. R. Astr. Soc.* **98**, 112.
ADAM, M. G., 1940. *Observatory*, **63**, 178; *Mon. Not. R. Astr. Soc.* **100**, 595.
BJERKNES, V., 1926. *Astrophys. J.* **64**, 93.
DE LURY, R., 1939. *J. R. Astr. Soc., Canada*, **33**, 245.
EVERSHED, J., 1936. *Mon. Not. R. Astr. Soc.* **96**, 152.
EVERSHED, J., 1938. *Mon. Not. R. Astr. Soc.* **98**, 195.
MCCREA, W. H. and MITRA, K., 1934. *Observatory*, **57**, 379.
ROYDS, T., 1937. *Mon. Not. R. Astr. Soc.* **97**, 692.
ST. JOHN, C. E., 1928. *Astrophys. J.* **67**, 195.
THACKERAY, A. D., 1937. *Mon. Not. R. Astr. Soc.* **97**, 672.
WESSELINK, A. J., 1940. *Bull. Astr. Instn. Netherlds.* no. 328.
WOOLLEY, R. v. D. R., 1937. *Mon. Not. R. Astr. Soc.* **98**, 3.

§ 5

ALLEN, C. W., 1937. *Astrophys. J.* **85**, 165.
CHANDRASEKHAR, S., 1933, 1934. *Mon. Not. R. Astr. Soc.* **94**, 14, 726.
DAVIDSON, C. R., MINNAERT, M., ORNSTEIN, L. S. and STRATTON, F. J. M., 1928. *Mon. Not. R. Astr. Soc.* **88**, 536.
MCCREA, W. H., 1929. *Mon. Not. R. Astr. Soc.* **89**, 718.
MCCREA, W. H., 1935. *Mon. Not. R. Astr. Soc.* **95**, 509.
MENZEL, D. H., 1931. *Publ. Lick Obs.* **17**.
MENZEL, D. H. and CILLIE, G. G., 1935. *Circ. Harv. Astr. Obs.* **410**.
MILNE, E. A., 1924. *Mon. Not. R. Astr. Soc.* **84**, 354.

MILNE, E. A., 1926. *Mon. Not. R. Astr. Soc.* **86**, 459.
MITCHELL, S. A., 1930. *Astrophys. J.* **71**, 1.
PANNEKOEK, A. and MINNAERT, M., 1928. *Verh. Akad. Wet. Amst.* **13**, no. 5.

§ 6

D'AZAMBUJA, L. *C.I.U.S. 5th Report, Solar-terr. Rel.* p. 45.
HALE, G. E., 1939. *Astrophys. J.* **70**, 265.
HULME, H. R., 1939. *Mon. Not. R. Astr. Soc.* **94**, 634.
KEENAN, P. C., 1935. *Astrophys. J.* **82**, 369.
LYOT, B., 1939. *Mon. Not. R. Astr. Soc.* **99**, 580.
MCMATH, R. R., 1937. *Publ. Michigan Obs.* **7**, no. 1.
MCMATH, R. R. and SAWYER, H. E., 1939. *Publ. Michigan Obs.* **7**, no. 9.
PETTIT, E., 1932. *Astrophys. J.* **75**, 9.
ZANSTRA, H., 1939. *Mon. Not. R. Astr. Soc.* **99**, 499.

§ 7

D'AZAMBUJA, L., 1939. *Bull. Astr., Paris*, **11**, 349.
ALLEN, C. W., 1940. *Observatory*, **63**, 93; *Mon. Not. R. Astr. Soc.* **100**, 635.
APPLETON, E. V., 1939. *C.I.U.S. 5th Report, Solar-terr. Rel.* p. 57.
ARCHENHOLD, G. H., 1939. *Mon. Not. R. Astr. Soc.* **99**, 723.
DOBBIE, J. C., MOSS, W. and THACKERAY, A. D., 1938. *Mon. Not. R. Astr. Soc.* **98**, 606.
GIOVANELLI, R. G. and HIGGS, A. J., 1939. *Terr. Magn. Atmos. Elect.* **44**, 181.
GIOVANELLI, R. G., 1940. *Astroph. J.* **91**, 334.
MÖGEL, 1930. *Telefunkenztg.* **56**, 14.
NEWTON, H. W., 1934. *Mon. Not. R. Astr. Soc.* **94**, 472.
NEWTON, H. W., 1939. *Observatory*, **62**, 318.
NEWTON, H. W. and BARTON, H. J., 1937. *Mon. Not. R. Astr. Soc.* **97**, 594.
RICHARDSON, R. S., 1935. *Publ. Astr. Soc. Pacif.* **47**, 325.
RICHARDSON, R. S., 1937. *Publ. Astr. Soc. Pacif.* **49**, 82.
RICHARDSON, R. S. and MINKOWSKI, R., 1939. *Astrophys. J.* **89**, 347.
WALDMEIER, M., 1938. *Z. Astrophys.* **16**, 276.

§ 8

ADAMS, W. S. and JOY, A. H., 1934. *Publ. Astr. Soc. Pacif.* **46**, 223.
BABCOCK, H. D. and H. W., 1934. *Publ. Astr. Soc. Pacif.* **45**, 249.
DUFAY, J. and GROUILLER, H., 1933. *C.R. Acad. Sci., Paris*, **196**, 1574.
DUNHAM, T., 1939. *Techn. Contr. Na. Geolog. Soc.* no. 1.
DYSON, F. W. and WOOLLEY, R. v. D. R., 1937. *Eclipses of the Sun and Moon.*
EVERSHED, J., 1939. *Mon. Not. R. Astr. Soc.* **99**, 217.
GOUDSMIT, S. and WU, T. Y., 1934. *Astrophys. J.* **80**, 154.
GROTRIAN, W., 1934. *Z. Astrophys.* **8**, 124.
JOHNSON, J. J., 1934. *Publ. Astr. Soc. Pacif.* **46**, 226.
KIEPENHEUER, K. O., 1935. *Z. Astrophys.* **10**, 260, 279.
LYOT, B., 1931. *Astronomie*, **45**, 248.
LYTTLETON, R. A. and HOYLE, F., 1939. *Proc. Camb. Phil. Soc.* **35**, 592.
MINNAERT, M., 1930. *Z. Astrophys.* **1**, 209.
MOORE, J. H., 1933. *Publ. Astr. Soc. Pacif.* **45**, 147.
ROSSELAND, S., 1933. *Publ. Oslo Obs.* no. 5.
SKELLETT, A. M., 1940. *Bell Syst. Tech. J.* **19**, 249.
STRATTON, F. J. M., 1937. *Nature, Lond.*, **140**, 682.
WALDMEIER, M., 1939. *Z. Astrophys.* **19**, 21.
WILDT, R., 1939. *Astrophys. J.* **89**, 295.
YOUNG, R. K., 1911. *Bull. Lick Obs.* **6**, 166.

SURFACE TENSION

By R. C. BROWN, Ph.D.,
University College, London

§1. THE PHYSICAL INTERPRETATION OF GIBBS'S ADSORPTION EQUATION

There have recently been several attempts to clarify the physical significance of Gibbs's adsorption equation, and, although the following is by no means an exhaustive account of all contributions to the subject, it is hoped that the reader may gain from it some idea of the problems involved.

In defining Γ, the "surface excess per unit area" of one of the components of a liquid mixture, Gibbs made use of a hypothetical mathematical plane which was supposed to separate one homogeneous phase from the other. The numerical value of Γ for any particular component depends upon the location of this plane. A common procedure in the case of a binary mixture is to imagine the plane drawn in such a position that the surface excess of one component is zero. The Gibbs equation then gives the particular value of the surface excess of the other component, which corresponds to this condition, in terms of the variation of the surface tension of the mixture with other properties of this component in the mixture. It is, however, difficult to visualize this position of the plane with respect to a liquid mixture in contact with its vapour, and, moreover, it is fairly obvious that the surface properties of such a system will depend upon the surface concentrations of both the components, so that the method of calculation which gives the surface excess of one component subject to the condition that that of the other is zero is not, perhaps, the most satisfactory from the physical point of view. In other words, it would be more useful to isolate that region of the system which contains the surface- or transition-layer between liquid and vapour and to calculate the excess concentrations of both components in this region. Something on these lines has been worked out by Guggenheim (1940) and will be referred to later.

Guggenheim and Adam (1932) went into the question of the definition of Γ by reference to the single mathematical plane, and their paper contains a systematic exposition of the various possible definitions, their inter-relationship, and the form of the Gibbs equation to which each leads.

The authors showed that in the simple derivation of the equation which gives it in the form

$$\Gamma_2 = c_2 \frac{\partial \gamma}{\partial \Pi},$$

where Γ_2 is the surface excess of component 2 whose concentration in a binary mixture is c_2, and whose osmotic pressure is Π, it is implied that the surface

excess is defined as being the number of moles of substance 2 in a portion of liquid containing unit area of surface *minus* the number of moles of 2 in an exactly equal volume in the interior. On this convention the geometrical surface coincides with the physical boundary of the liquid. The ratio of Γ_2, defined by reference to the position of the surface for which Γ_1 is zero, to Γ_2 defined as above is equal to $1+\frac{N_2V_2}{N_1V_1}$, where N_1 and N_2 are molar fractions and V_1 and V_2 are partial molar volumes. Thus the two definitions give identical values of Γ_2 in the immediate neighbourhood of $N_2=0$.

Guggenheim and Adam compared the numerical values of Γ_2 when defined in the various ways in the case of the system: water-ethyl, alcohol. They also suggested that the surface excess of alcohol in this system might be defined on the assumption that the interface consisted of a monomolecular layer. The Gibbs mathematical surface then separated the homogeneous liquid from the surface layer, and the surface excesses of each component were the actual surface concentrations per unit area. These could not be calculated unless some simultaneous relationship expressing this particular method of defining Γ were assumed. Guggenheim and Adam suggested that for alcohol and water such a relationship might take the form of an expression of the areas occupied separately by the molecules in the surface. Thus they wrote

$$A_1\Gamma_1+A_2\Gamma_2=1,$$

where A_1 and A_2 were the areas occupied in a monomolecular layer by one mole of water and alcohol respectively. A_2 could be assigned a value without much difficulty, assuming that the alcohol molecules were oriented perpendicular to the surface; but A_1 was less definite, and the authors took, in turn, three possible values of it, viz., $A_2/3$, $A_2/2$ and $2A_2/3$. The value of A_1 was found to affect the trend of the curves of surface excess *vs.* concentration but little, and the actual figures obtained were in such reasonable agreement with others derived without reference to the monomolecular surface-structure that the assumption might be said to be quite a possible one. Figures for $\Gamma_2/(\Gamma_1+\Gamma_2)$ showed a steady rise in going from pure water to pure alcohol except over the region 0·2 to 0·4 molar fraction of alcohol, where there was a slight downward slope. This unlikely change of slope was ascribed by Guggenheim and Adam to a variation of A_1 and A_2 with composition which was not taken into account. At any rate they preferred this explanation to that of Schofield and Rideal (1925), who suggested that at molar fractions of alcohol above 0·3 the surface was a complete monomolecular alcohol film resting on a layer containing excess of water.

More recently van Rysselberghe (1938) has suggested the use of a simple mixture rule in the place of Guggenheim and Adam's equation $A_1\Gamma_1+A_2\Gamma_2=1$. The rule expresses in the simplest possible manner the dependence of the surface tension of the mixture upon surface concentration, thus

$$\gamma=\frac{\Gamma_1\gamma_1+\Gamma_2\gamma_2}{\Gamma_1+\Gamma_2}$$

This combined with Gibbs's equation in the form

$$d\gamma = (\Gamma_1 N_2/N_1 - \Gamma_2)\, d\mu_2,$$

where μ_2 is the molar chemical potential of the alcohol, will give numerical values of Γ_1 and Γ_2. Van Rysselberghe takes the values of $d\gamma/d\mu_2$ from Guggenheim and Adam's paper and finds that as N_2 increases from 0 to 1, $\Gamma_2/(\Gamma_1+\Gamma_2)$ increases continuously from 0 to 1 *without* showing a negative slope over any concentration range. He is thus inclined to consider his method of calculation to be more acceptable than that of Guggenheim and Adam. The values which van Rysselberghe obtains for the area of the surface occupied by the alcohol molecules suggest that the latter are oriented parallel to the surface in pure alcohol and perpendicular to the surface in dilute solutions. Van Rysselberghe notes that the distance between his dividing plane (which is immediately below the monomolecular layer) and the plane corresponding to the convention that the surface excess of water is zero ($\Gamma_1=0$) varies from 1·09 A. for $N_2=0{\cdot}05$ to 0·16 A. for $N_2=0{\cdot}9$. The "$\Gamma_1=0$" convention, of course, gives no information about the surface concentration of water.

Guggenheim (1940) has recently attempted to bring the definition of surface excess more into line with the modern physical picture of a liquid separated from its vapour by a surface film or layer of small but finite thickness and having properties peculiar to itself. The clarity and usefulness of any exposition of the thermodynamics of an interface depends to a large extent upon the definitions of the quantities involved, and it is here that Guggenheim is careful to adhere to physical realities. He gives all his symbols a simple physical definition by considering *two* planes parallel to the liquid surface, one in the vapour and the other in the liquid. These planes separate the interface from the bulk phases, but for the purpose of the analysis it is not necessary for them to be coincident with its boundaries. Adsorption equations appropriate to this method of definition are worked out.

Guggenheim believes that "the mathematical elegance of Gibbs's formulae is attained only at the cost of a certain obscurity in the physical significance of the symbols used", and he remarks "Experience indicates that it is much less difficult to use Gibbs's formulae than to understand them". This is a statement with which many who scan the journals devoted to physical chemistry will be inclined to agree. It may well be that a clarification of the definition of the term "surface excess" and a clearer appreciation of the relevance or unsuitability of the "$\Gamma_1=0$" convention to a particular problem will lead to a smoothing out of some of the present difficulties with regard to the interpretation of results and, possibly, also to a modification of some of the published adsorption data.

§ 2. ADSORPTION AND THE SURFACE TENSION OF SOLUTIONS

In an earlier volume of these *Reports* McBain and Mills (1938) gave a calculation of the adsorption of lauryl sulphonic acid in the surface of aqueous solutions using Gibbs's equation together with surface-tension and osmotic-pressure data. They adopted the usual "$\Gamma_1=0$" convention referred to in § 1 above. Now,

the curve of surface tension *vs.* concentration of lauryl sulphonic acid in water is of McBain's Type III, i.e. it shows a marked fall of surface tension at low concentrations followed by a slight rise and an approach to a steady value as concentration is further increased. It is inevitable that the Gibbs equation should yield values for the surface concentration of the acid which are difficult to interpret. Thus, where the curve of surface tension *vs.* concentration passes through a minimum ($dy/dc=0$) the equation indicates zero adsorption, although the actual surface tension may be 30 or 40 dynes/cm. below that of pure water. Beyond the minimum the equation must suggest a deficiency of acid in the surface—again a difficult conception when the surface tension is much lower than that of water.

In a more recent paper McBain and Wood (1940) publish the results of surface-tension measurements made by the du Noüy ring method on these solutions. The results confirm those obtained by other methods and show a decrease of surface tension with time in the case of solutions of concentrations less than 0·01 *N*. These authors have also investigated the surface adsorption by the very direct method of slicing the surface off with a microtome and examining its composition. This technique (referred to in vol. 4 of these *Reports*) has been brought to a high state of precision. Results show that positive adsorption of lauryl sulphonic acid takes place at all bulk concentrations, and there appears to be no measure of agreement whatever between these results, which, as the authors remark, are difficult to question on account of the precautions taken, and the figures given by McBain and Mills's (1938) method of calculation from Gibbs's equation. An interferometer method verifies that the adsorption is positive for all bulk concentrations.

Bikermann (1938) has put forward a theory of the surface tension of aqueous solutions of electrolytes. He expresses the surface tension as the sum of various terms, one of which, the "electrostatic energy inherent in the unit of the surface", is shown to decrease as the concentration of the electrolyte increases from zero. The difference between this term and the corresponding one for pure water (which, in fact, will always contain a few ions) is shown to reach a maximum value independent of the nature and amount of the electrolyte present. This affords an explanation of the results of Jones and Ray (1937), who found that the curves of surface tension *vs.* concentration for three inorganic salts fell slightly from the value for pure water to a minimum. The concentration at which the minimum occurred was the same for each salt. Langmuir (1938) is inclined, however, to ascribe the "Jones-Ray effect" to the action of a water film adhering to the wall of the capillary tube by reason of electrification of the latter.

The time-variation of the surface tension of some dilute solutions has been measured by Fisher and Mackay (1938). The method used was that of Ferguson and Hakes (1929), which consists in measuring the pressure required to keep the meniscus in equilibrium at the bottom of a capillary tube. They have published curves of surface tension *vs.* time for M/4000 sodium oleate and saturated and half-saturated solutions of lauric acid. The usual slow fall with time is noticed,

and a point of special interest is that, whereas the sodium oleate solution shows an initial rapid fall immediately the surface is formed, the concentrated lauric acid solution delays for about one minute before the rapid fall begins. This period of delay is shorter at higher temperatures. Ferguson and Sharp (1934) found a similar effect with acid solutions *and* with sodium oleate. Fisher and Mackay compare their results with the theory of Bond and Puls (1937) which ascribes the time effects to diffusion of dissolved molecules into and out of the surface. The equation to be tested is

$$\frac{(\gamma_0-\bar{\gamma})-(\gamma_0-\gamma)}{\gamma_0-\bar{\gamma}} = e^{-\frac{2}{\sqrt{\pi}}\sqrt{\frac{t}{T}}},$$

where γ_0 is the surface tension of water, γ is the surface tension of the solution t seconds after the new surface is formed, $\bar{\gamma}$ is the final static surface tension and T is the time taken for the surface tension to pass half-way to its static value. Fisher and Mackay write the equation in the form

$$\frac{\gamma_0-\gamma}{\gamma_0-\bar{\gamma}} = 1-e^{-\frac{2}{\sqrt{\pi}}\sqrt{\frac{t}{T}}}$$

and draw graphs of $(\gamma_0-\gamma)/(\gamma_0-\bar{\gamma})$ against t/T, choosing the values for T for each solution which give the best fit between the theoretical curve and the experimental points. The fit is successful only in the case of the half-saturated lauric acid solution.

A novel method of studying the surface structure of aqueous solutions of such positively adsorbed substances as caprylic acid, hydrocinnamic acid, and phenol has been devised by Ford and Wilson (1938). By cinematography they record the spreading of oleic acid on the liquid surface which has previously been dusted with talc. The adsorbed film of solute behaves in much the same manner as an insoluble film on a water surface. The progress of the spreading is governed by the properties and behaviour of both the oleic acid and the surface solution; but the main purpose of the experiment is to study the final stage of spreading in which the soluble surface film becomes "compact" owing to its high compression in the area between the oleic-acid film and the boundaries of the dish. At this stage the rate of reduction of area is comparatively low and may be assumed to depend only on the rate of escape of molecules from the soluble film. Here the curve of oleic-acid area *vs.* time is linear, and extrapolation of this portion to zero time will enable an estimate to be made of the area occupied by a single solute molecule when closely packed in the surface. These areas agree with values obtained by other methods. The surface concentration calculated from Gibbs's equation and surface-tension data is also in reasonable agreement with values found by extrapolation.

§3. CONTACT ANGLE AND WETTING

It is a well-known fact that the properties of a solid surface may be modified when it is exposed to the vapour of a liquid, on account of the formation upon it of an adsorbed layer of the liquid. Moreover, the properties of the adsorbed

layer differ from those of the liquid in bulk, especially if the liquid is strongly polar, so that a high degree of orientation of its molecules occurs. Bartell and Bristol (1940) have measured the contact angle in a liquid drop formed on a solid surface which has been exposed to water vapour at various pressures. They distinguish between dynamic advancing, dynamic receding and static advancing contact angles. The first and second are respectively those observed when liquid is being added to the drop and when it is being removed. The third angle is the one observed when the advance is halted. Ignoring these distinctions we can express Bartell and Bristol's main results as follows. The contact angle between an organic liquid and a water-attracting (hydrophilic) solid, such as quartz or Pyrex glass, is increased by previous exposure of the solid to water vapour, the higher the vapour pressure the greater the contact angle. Moreover, the contact angle in the organic liquid becomes practically independent of the nature of the solid if the latter is covered with an adsorbed film formed from saturated water vapour even though, in the absence of the water layer, the angles for the various solids may be very different.

The experiments were also carried out the other way about: organophilic solid surfaces were exposed to organic vapours and measurements were made of the contact angle in a water drop on the surface. It sometimes happened in this case that the receding angle became zero, indicating that the advancing water drop had swept the organic film off the solid surface.

Taking the case of an organic liquid (γ_L) on a surface completely covered with a water film (γ_W') and making the assumption that interfacial tension (γ_{LW}) between the water and the drop is the same as that between the two liquids in bulk (a questionable assumption according to the authors), we can write

$$\gamma_W' = \gamma_{LW} + \gamma_L \cos\theta,$$

where θ is the static advancing angle. Thus γ_W' may be determined and is found to be 77 dynes/cm. Similar calculations carried out for adsorbed organic films indicate that the surface tension of such films is greater than the bulk value in the case of polar liquids, but that there is little difference with non-polar liquids.

Bikerman (1940) has made some contact-angle measurements with liquids resting on "built-up" multilayers. He bases his method on the following interesting observation. "If a water drop be put on a multilayer deposited on a chromium-plated slide it changes form within a few seconds. A fraction of a second after it has touched the surface the contact angle viewed with a projection microscope may be, say, 90°, but it diminishes rapidly." A typical example shows a fall of contact angle from 90° to 42° in 30 minutes. When the drop is then removed from the slide it leaves on the multilayer a coloured trace in the form of a ring—a case of what is known as water-line corrosion, the progress of which is responsible for the decreasing contact angle. The diameter of the ring left by a drop corresponds to the initial contact angle, the value of which can be found from the volume of the drop and the diameter of the ring. The method may also be used in cases where the liquid evaporates without leaving a mark

if the drop is sprayed with ignited talc which slides down to the circumference of the drop.

In discussing wettability Devaux (1940) points out that a thick film of liquid on a solid surface is not stable. Water poured on to a freshly formed glass surface breaks up into drops which are separated by a monomolecular layer of water adhering strongly to the glass surface. On the other hand, to take a case of non-wetting, if water is poured on to a wax surface, the area not occupied by drops is not covered with a monomolecular layer, although the drops themselves are held fixed to the surface by adhesion. Devaux suggests therefore that good and bad "wetting" are characterized respectively by the formation or absence of the monomolecular film—conditions which in turn are indicated by the time taken for the formation of drops after wetting. Thus, rapidly formed drops represent poor wetting. Devaux further emphasizes the fact that the hydrophobic or hydrophilic nature of the surface of a particular substance depends very much upon the mode of preparation of the surface. Its nature is not likely to be uniform over large areas: in some places the water-attracting ends of the molecules may be exposed and in others they may be hidden. He explains the adhesion of water drops to a window pane by suggesting that they attach themselves to regions where the surface is hydrophilic.

To a symposium on surface-active agents organized by the Divisions of Industrial and Engineering Chemistry and Colloid Chemistry of the American Chemical Society, Hauser, Andreas and Tucker (1939) contribute a survey of the various methods of measuring surface tension, with particular reference to static methods capable of giving reliable data about the change of surface tension with time. They are inclined to favour the pendant-drop method (referred to later in this article) for this purpose. Incidentally, the account which these authors give of the practical applications to which surface-tension measurements are now put would surprise many a physicist whose interests lie in other directions. In the same symposium Caryl and Ericks (1939) discuss the action and chemical structure of wetting agents, i.e. substances which, when dissolved in water, increase its tendency to wet a given solid surface. In particular they describe the performance of one such agent, Aerosol OT Dry. So magnificent is this that it will overcome the proverbial non-wettability of the duck. Stirring photographs are reproduced of the rescue of a duck almost completely engulfed by a dilute solution of this substance.

§4. METHODS OF MEASUREMENT OF SURFACE TENSION AND RESULTS

In searching for a method of examining the change of surface tension with time, Andreas, Hauser and Tucker (1938) have chosen that depending on the measurement of the dimensions of pendant drops. Earlier work on this method of measuring surface tension has involved the optical projection of an image of a drop hanging from the lower end of a vertical tube. The outline of the image

was then traced, and surface tension could be calculated from the shape of the curve. The method was used by Worthington (1881), and its possibilities were further investigated by Ferguson (1912). Andreas, Hauser and Tucker take photographs of the drops by a method which eliminates distortion of the outline and calculate the surface tension from the following equation:

$$\gamma = \frac{g(\rho_1 - \rho_2)\, d_e^2}{H},$$

where ρ_1 is the density of the liquid, ρ_2 that of the surrounding gas or liquid, d_e is the maximum diameter of the drop, and H is a function whose value depends on the shape (S) of the drop. S is defined as d_s/d_e, where d_s is the diameter of the drop at a plane whose height above the lowest point of the drop is d_e. The authors have drawn up a table showing the variation of $1/H$ with S, the figures having been arrived at by making measurements on photographs of water drops of various sizes. They claim an uncertainty of $\pm 0{\cdot}5$ per cent in surface-tension values determined by this method, and 1 c.c. of liquid is sufficient for several determinations. The authors' measurements of the change of surface tension with time in the cases of dilute solutions of sodium oleate and sodium stearate in air and also against mineral oil show the following main features: (*a*) for very dilute solutions (0·00025 per cent) a slight raising of the surface tension of water, (*b*) a greater rate of fall of surface tension with time for higher concentrations, (*c*) the existence of a minimum in the curves of surface tension *vs.* time at certain concentrations.

Bosworth (1938) has examined the applicability of Jaeger's maximum bubble-pressure method to the measurement of the surface tension of mercury and has come to the same conclusions as R. C. Brown (1928). The method is workable and satisfactory but for the fact that, when jets of, say, glass or silica are used, the radius of the bubble which exhibits the maximum pressure is liable to vary with successive bubbles and, even if it does remain constant, one cannot be certain whether the bubble detaches itself from the inner or outer circumference of the jet. The trouble is due entirely to the smaller adhesion for the material of the jet exhibited by mercury as compared with the usual wetting liquids upon which the method is more frequently employed with such success. Bosworth overcomes the difficulty by using an amalgamated copper jet, the internal and external radii of which are practically identical. He has used the method to study the change of the surface tension of mercury with temperature when the bubbles are blown with air containing various partial pressures of acetic acid vapour, and has obtained an interesting series of curves (Bosworth, 1939*a*). In all cases the values are lower than those obtained with pure air, the diminution increasing with the partial pressure of the acid vapour. The temperature coefficient is also affected, being in no case constant throughout the temperature range. When greater proportions of vapour are present, the curves of surface tension *vs.* temperature show a maximum. By plotting the amount of absorption at any particular temperature

against the partial pressure of the acetic-acid vapour it is shown that the adsorption increases with pressure and diminishes with increasing temperature.

Bosworth (1939 b) has also investigated the effect on the surface tension of mercury of the presence of carbon dioxide in the air with which the bubbles are blown. He finds that "with increasing content of carbon dioxide the surface tension at first falls to a minimum some 8 per cent lower than the value in pure air, at about 3 per cent carbon dioxide, and thereafter rises very slowly with increasing carbon-dioxide content, the value in pure carbon dioxide being about $1\frac{1}{2}$ per cent, lower than the value obtained in pure air".

Bate (1939) has described a method of estimating the surface tension of a sample of mercury *in vacuo* in a sealed glass tube which happened to show an angle of contact of 90°. By setting the tube in a horizontal position he obtained, as it were, a slice of a large sessile drop, and since the contact angle was 90° the base of the drop was also the section of largest area. Thus measurement of the distance of the highest point of the drop above the base would enable the surface tension to be calculated. This height was arrived at by extrapolation from measurements made on the mercury surface when the tube was slightly tilted. A value of 490 dynes/cm. was obtained.

Using the method in which a liquid thread in a horizontal capillary tube is pushed to the end and the exposed surface made plane, Cockett and Ferguson (1939) have measured the surface tensions of water and heavy water in the temperature range 15° c. to 75° c. A rough and ready method of stating their results is to say that over this temperature range the surface tension of heavy water at T° c. is equal to that of ordinary water at $(T+7{\cdot}5)$° c. Applying the Eötvös equation,

$$\frac{d}{dt}\left[\gamma\left(\frac{M}{\rho}\right)^{2/3}\right] = -2{\cdot}1,$$

they find that between 5° c. and 40° c. the effective molecular weight of ordinary water is 3·4 times its actual molecular weight, while for heavy water the factor is 2·8. Thus it appears that heavy water is less associated than ordinary water.

Katz and Saltman (1939) have found that a single continuous curve is obtained if the surface tensions of a series of hydrocarbons such as the normal paraffins are plotted against their reduced temperatures (i.e. the absolute temperature at which the surface-tension measurements are made divided by the absolute critical temperature of the liquid). This leads the authors to suggest that a surface-tension measurement made at room temperature on a particular mixture might yield an estimate of its critical temperature when reference is made to the curve and the corresponding reduced temperature read off. This involves the necessity that both the surface tension and the critical temperature of the mixtures should have approximately additive properties.

Bradley (1939), using a simple film method, has studied the elevation of the surface tension of paraffin caused by finely divided lamp-black and other powders. A 32 per cent suspension of carbon in paraffin has a surface tension 52 dynes/cm.

greater than that of the pure oil. The author has worked out a theory of the effect in terms of an inverse seventh-power law of attraction.

Lemonde (1938) has designed a tensiometer for the measurement of the pull on a ring or frame as it is detached from a liquid surface. He replaces the torsion balances of Searle and du Noüy by a steel blade supported cantilever-fashion. Movements of the free end of the blade are measured by a microscope and a damping device is fitted. Lemonde has used the instrument to investigate how the downward pull on a vertical frame or horizontal ring varies with the height to which it is raised above the free surface of the liquid. He verifies the well-known fact that, as the frame or ring is steadily raised, the pull rises to a maximum and then diminishes. Accurate surface-tension values are difficult to obtain by these methods, of course, owing to the uncertainty of the shapes of the liquid films, and, especially in the case of the ring, the weight of the raised liquid. Correction factors have been evaluated for the ring method (Harkins and Jordan, 1930).

Tyler (1938 and 1939) has introduced the straight-line graph into two common surface-tension experiments. The first of these is Wilhelmy's method of measurement. A vertical glass plate is suspended from the arm of a torsion balance and its apparent loss of weight (m) is measured for different depths of immersion (h). The graph of m against h is a straight line from which, together with the dimensions of the plate, the surface tension and density can be calculated. The author points out that the glass plate may be replaced by a vertical glass tube or rod. The second method is applicable only to liquids which form films. When a soap film is formed between two vertical threads which join the extremities of two horizontal rods, the lower one of which is hanging freely, the threads assume circular shapes. An image of the system is projected on to a screen and the variation of the dimensions of the film with the combined weight of the lower rod and any load which it may carry is observed. A straight-line graph may be drawn and surface tension calculated. Both this and the glass-plate experiment serve to illustrate the use of the straight-line graph and to emphasize the theory of the methods, even though the experimental procedure is rather more complicated than that required for single determinations of surface tension.

§ 5. MISCELLANEOUS RESEARCHES

Theoretical. A theory of surface tension modelled on Andrade's (1934) theory of viscosity has been put forward by Sibaiya and Rama Rao (1939). They assume that at the melting point the molecules of a substance continue to oscillate with the Lindemann frequency about a slowly displaced equilibrium position, and also that the surface tension of a liquid may be defined as being equal to the total energy of the molecules in a surface layer of unit area and unimolecular thickness. The following expression is then obtained for γ:—

$$\gamma = \frac{2\pi a^2 m \nu^2}{\sigma^2},$$

so that $$\nu = k\sqrt{\gamma/m}$$

where σ = mean distance between two molecular centres,
a = amplitude of vibration,
ν = Lindemann frequency,
m = mass of atom or molecule.

The authors then assume that a/σ is a constant and that

$$\nu = 2{\cdot}8 \times 10^{12}\sqrt{(T_s/MV^{1/3})},$$

so that

$$k = 2{\cdot}8 \times 10^{12}\sqrt{\frac{T_s}{MV^{1/3}}}\Big/\sqrt{\gamma/m},$$

where T_s = melting point,
M = atomic or molecular weight,
V = atomic or molecular volume.

Taking data for 45 elements and compounds the authors tabulate the values of k. These are found to range from 0·91 to 3·68 with a "mean value" of 2·34.

In order to represent the variation of surface tension with temperature Sibaiya and Rao use an equation due to Macleod (1938) which expresses the variation of molecular frequency with temperature.

Thus

$$\nu_T = \nu\frac{v_s - v_0}{v_T - v_0}\sqrt{\frac{T}{T_s}},$$

where ν_T is the frequency at an absolute temperature T, v_s the specific volume of the substance at the melting point, v_T the specific volume at $T°$K. and v_0 is a constant. From this equation and the authors' expression for γ quoted above it follows that

$$\gamma_T \propto \frac{T}{(v_T - v_0)^2}.$$

If the energy of a molecule is inversely proportional to the mth power of the distance between two molecular centres, the authors find evidence for ascribing to m the value 10. Thus if the molecular energy varies as σ^{-m} then $\gamma \propto 1/\sigma^{m+2}$, and, further, $\sigma \propto V^{1/3}$, where V is the molar volume of the liquid, so that the expression $\gamma V^{(m+2)/3}$ must be a constant. This reduces to the parachor if $m = 10$. Again, if the last expression is combined with the Eötvös equation, it follows that $V^{m/3}(T_c - T)$ is also a constant which leads to an expression for the coefficient of cubical expansion at any temperature, T, namely, $\alpha_T = 3/\{m(T_c - T)\}$, which agrees with experiment if $m = 10$.

Finally, by combining the Eötvös equation with $\gamma V^{(m+2)/3} =$ constant it is found that $\gamma \propto (T_c - T)^{(m+2)/3}$. Experimentally the index is found to be 1·2, which again suggests that $m = 10$.

The authors have also adapted their theory to give an expression for the

viscosity of a liquid. Taking the mean value of k obtained from surface-tension data, namely, 2·34, they find that

$$\eta = \frac{4}{3 \cdot 3} \frac{mv}{\sigma}.$$

Andrade's expression is

$$\eta = \frac{4}{3} \frac{mv}{\sigma}.$$

An empirical relation between surface tension and viscosity has been recorded by Buehler (1938). He quotes a relation, due to Mott Souders (1938), viz.: $I = M\{\log_{10}(\log_{10}\eta) + 2 \cdot 9\}/\rho$, where I is the viscosity-constitutional constant, M the molecular weight, ρ the density and η the viscosity. The relation is the "viscosity" counterpart of the parachor expression $P = M\gamma^{1/4}/\rho$. Buehler finds from experimental data that an average value for I/P is 1·2, and this leads to the following relation between η and γ:—

$$\log_{10}(\log_{10}\eta) = 1 \cdot 2\gamma^{1/4} - 2 \cdot 9.$$

Brillouin (1938) has called attention to the fact that two definitions of surface tension are in existence. The first is really the definition of free surface energy. It applies to pure liquids, and is the one with which the classical theories of surface tension are concerned. Thus it is the work done when a liquid surface is increased by unit area under conditions in which the structure of the surface remains constant, i.e. the new area is occupied by molecules drawn from the interior of the liquid. The separation of the surface molecules remains constant. This definition will apply to solutions in which molecules of solvent and solute diffuse rapidly to the surface which is being extended. The total free energy per unit area of new surface is the sum of the separate energies gained by the molecules of both solvent and solute in being drawn into the surface.

In the second definition of surface tension we speak of the force per unit length acting on a free rod which forms the fourth side of a rectangular frame containing a surface but no liquid in bulk. When the surface is extended under these conditions its structure undergoes a change. The total number of molecules constituting the surface remains the same but their separation increases with the extension. Thus we have the exact analogue of a stretched rubber membrane and also of the insoluble monomolecular film of the Langmuir-Adam experiments. The surface tension defined thus varies with the area of the surface.

Mahajan (1939) has worked out how the life of a drop floating on the surface of its own liquid depends on the viscosity of the liquid. Density and surface tension do not influence the life of the drop.

Experimental. Littaye (1939) has studied one or two points in connexion with liquid jets. The constrictions on a jet which precede its breaking up into drops divide the jet into sections which can act as sphero-cylindrical lenses and produce focal lines at right angles to the axis of the jet. The shadow of the jet cast on a distant screen is traversed by bright lines even when the constriction

cannot be seen by direct examination of the jet itself. The maximum and minimum diameters of the jet may also be deduced.

If the jet is oscillating the focal lines will not be perpendicular to its axis, and so the presence of small oscillations on the jet may be detected. It is found that even if the flow is non-turbulent, oscillations may be present, and Littaye further shows that the origin of the oscillations is in the liquid which wets the end of the tube. When this wetting is prevented by smearing grease on the end of the tube, oscillations are not found on the jet unless the flow in the tube is turbulent.

A small drop of water evaporates in saturated water-vapour on account of its higher vapour pressure. If the drop contains dissolved substance, however, its vapour pressure is thereby lowered, and for a given concentration there should be a definite drop-radius which gives equilibrium. This condition is represented by the equation

$$\frac{2\gamma}{r\rho}\frac{M}{RT}p_0 = np_0,$$

where r is the radius of the drop, ρ the density of the solution, M the molecular weight and p_0 the vapour pressure of the solvent, R the gas constant, T the absolute temperature, and n is the concentration expressed as $n_s(1+\delta)/(n_s+n_f)$, i.e. n_s gm. molecules of solute in n_f gm. molecules of solvent, δ being the degree of dissociation. Rumpf and Geigl (1938) have tested this relationship by measuring the radii of "equilibrium drops" of aqueous solutions in a stream of saturated water vapour. The results show that, although the equation is not quantitatively obeyed, the size of the equilibrium drop depends only on concentration. Thus, if the equation fails, calculations of the concentration of dissolved substance in fog droplets must be suspect.

Narasimha Murti and Seshadri (1939) have made a large number of observations of the movements of organic solids on the surface of water and have attempted to classify their behaviour. They divide those substances which show movements into three groups according to whether their movements are "vigorous", "slow" or "indifferent". As the authors remark, the necessary factors are non-uniform lowering of surface tension around the periphery of the fragment and the subsequent disappearance of the dissolved substance from the surface either by solution or by evaporation. Apparently, vigorous movements are shown by solids which lower the surface tension of water (in saturated solution) by as little as 3 per cent. The percentage lowering by camphor is 25·8, and phenylacetic acid, which lowers the surface tension of water by 32·5 per cent, performs the "most spectacular" movements of all. The authors discuss their classification in relation to the chemical constitution of the solids, and conclude by describing some interesting experiments on "combination movements" of pairs of substances, which should make successful lecture or conversazione experiments. The writer was interested to note that acetylsalicylic acid was in the "vigorous" group and promptly verified the fact by means of an aspirin tablet and a *clean* water surface.

An interesting instance of capillary action is reported by Afzal and Vaidhianathan (1939). It has been observed in parts of India that "when there has been a slight shower of rain or even a small amount of irrigation over a wide area, some of the wells, especially the shallow ones, rise out of all proportion to the amount of rain or irrigation". Afzal and Vaidhianathan have found an explanation of this occurrence and have demonstrated the effect by means of a model. A large tank has two vertical side tubes with strainers fitted near the bottom and is filled with sand and water. After the sand has been completely packed the water is drained off until it stands about half-way up the side tubes. If the tank is now allowed to stand so that water evaporates from the surface the level in the side tubes sinks by about 10 cm.—a fall which cannot be accounted for to any extent by evaporation. About 100 gm. of water are now sprinkled on the surface of the sand, and it is computed that this would cause a rise of a few tenths of a mm. in the side tube. Actually the rise is about 8 cm. The explanation is that as soon as evaporation occurs a large number of menisci are formed at the surface of the sand, and their curvature is such as to cause the pressure inside the tank to diminish by about 10 cm. of water. The sprinkling of water on the top of the sand restores the plane water surface.

REFERENCES

AFZAL, M. and VAIDHIANATHAN, V. I., 1939. *Proc. Indian Acad. Sci., Bangalore*, **A, 9**, 309.
ANDRADE, E. N. DA C., 1934. *Phil. Mag.* **17**, 497.
ANDREAS, J. M., HAUSER, E. A. and TUCKER, W. B., 1938. *J. Phys. Chem.* **42**, 1001.
BARTELL, F. E. and BRISTOL, K. E., 1940. *J. Phys. Chem.* **44**, 86.
BATE, A. E., 1939. *Phil. Mag.* **28**, 252.
BIKERMAN, J. J., 1938. *Trans. Faraday Soc.* **34**, 1268.
BIKERMAN, J. J., 1940. *Trans. Faraday Soc.* **36**, 412.
BOND, W. N. and PULS, H. O., 1937. *Phil. Mag.* **24**, 864.
BOSWORTH, R. C. L., 1938. *Trans. Faraday Soc.* **34**, 1501.
BOSWORTH, R. C. L., 1939 a. *Trans. Faraday Soc.* **35**, 1349.
BOSWORTH, R. C. L., 1939 b. *Trans. Faraday Soc.* **35**, 1353.
BRADLEY, R. S., 1939. *J. Phys. Chem.* **43**, 663.
BRILLOUIN, L., 1938. *J. Phys. Radium*, **9**, 426.
BROWN, R. C., 1928. *Phil. Mag.* **6**, 1044.
BUEHLER, C. A., 1938. *J. Phys. Chem.* **42**, 1207.
CARYL, C. R. and ERICKS, W. P., 1939. *Industr. Engng Chem.* (Indust. Ed.), **31**, 44.
COCKETT, A. H. and FERGUSON, A., 1939. *Phil. Mag.* **28**, 685.
DEVAUX, H., 1940. *C.R. Acad. Sci., Paris*, **210**, 27.
FERGUSON, A., 1912. *Phil. Mag.* **23**, 417.
FERGUSON, A. and HAKES, J. A., 1929. *Proc. Phys. Soc.* **41**, 214.
FERGUSON, A. and SHARP, L. E., 1934. *Rep. Progr. Phys.* **1**, 41.
FISHER, W. C. and MACKAY, C. A., 1938. *Canad. J. Res.* A, **16**, 207.
FORD, T. F. and WILSON, D. A., 1938. *J. Phys. Chem.* **42**, 1051.
GUGGENHEIM, E. A., 1940. *Trans. Faraday Soc.* **36**, 397.
GUGGENHEIM, E. A. and ADAM, N. K., 1932. *Proc. Roy. Soc.* A, **139**, 218.
HARKINS, W. D. and JORDAN, H. F., 1930. *J. Amer. Chem. Soc.* **52**, 1751.
HAUSER, E. A., ANDREAS, J. M. and TUCKER, W. B., 1939. *Industr. Engng Chem.* (Indust. Ed.), **31**, 32.
JONES, G. and RAY, W. A., 1937. *J. Amer. Chem. Soc.* **59**, 187.

KATZ, D. L., and SALTMAN, W., 1939. *Industr. Engng Chem.* (Indust. Ed.), **31**, 91.
LANGMUIR, I., 1938. *Science*, **88**, 430.
LEMONDE, H., 1938. *J. Phys. Radium*, **9**, 505.
LITTAYE, G., 1939. *C.R. Acad. Sci., Paris*, **208**, 1705.
MCBAIN, J. W. and MILLS, G. F., 1938. *Rep. Progr. Phys.* **5**, 30.
MCBAIN, J. W., and WOOD, L. A., 1940. *Proc. Roy. Soc.* A, **174**, 286.
MACLEOD, D. B., 1938. *Proc. Phys. Soc.* **50**, 788.
MAHAJAN, L. D., 1939. *Ind. J. Phys.* **13**, 299.
NARASIMHA MURTI, G. V. L. and SESHADRI, T. R., 1939. *Proc. Indian Acad. Sci., Bangalore*, A, **9**, 10.
RUMPF, E. and GEIGL, E., 1938. *Z. Phys.* **111**, 301.
VAN RYSSELBERGHE, P., 1938. *J. Phys. Chem.* **42**, 1021.
SCHOFIELD, R. K. and RIDEAL, E. K., 1925. *Proc. Roy. Soc.* A, **109**, 61.
SIBAIYA, L. and RAMA RAO, M., 1939. *Ind. J. Phys.* **13**, 293.
SOUDERS, MOTT, Jr., 1938. *J. Amer. Chem. Soc.* **60**, 154.
TYLER, E., 1938. *Sch. Sci. Rev.* **20**, 41.
TYLER, E., 1939. *Sch. Sci. Rev.* **20**, 389.
WORTHINGTON, A. M., 1881. *Proc. Roy. Soc.* A, **32**, 362.

EQUATIONS OF STATE *

By JAMES A. BEATTIE and WALTER H. STOCKMAYER

§1. INTRODUCTION

A complete treatment of the thermodynamic properties of real gases and mixtures of real gases should express these properties in terms of elementary functions and be suitable for the solution of problems involving chemical as well as physical changes in state. Such a treatment involves: (1) a general thermodynamic theory of gases and gas mixtures which states the results in terms of differential equations or definite integrals; (2) a statistical theory of gases and gas mixtures or an equivalent method (graphical or empirical) that gives equations which permit evaluation of the integrals, resulting from the thermodynamic method, in terms of certain parameters, or constants characteristic of the particular circumstances involved; and (3) the determination of the numerical values of parameters or constants for the substances involved.

A general thermodynamic theory for pure gases has long been known, and one for gas mixtures has been given (Gillespie, 1925, 1926 and 1936; de Donder, 1925; v. Lerberghe, 1925 and 1928; Beattie, 1928 a, 1928 b and 1930 a; de Donder and v. Rysselberghe, 1936). As examples of the results obtained the following equations may be cited:—

For pure gases:

Work content:

$$F_{VT}=\int_V^\infty\left[p-\frac{nRT}{V}\right]dV-nRT\ln\frac{V}{nRT}+nf'. \qquad \ldots\ldots(1)$$

Free energy:

$$F_{pT}=\int_0^p\left[V-\frac{nRT}{p}\right]dp+nRT\ln p+nRT+nf'. \qquad \ldots\ldots(2)$$

Fugacity:

$$RT\ln f=\int_0^p\left[\frac{V}{n}-\frac{RT}{p}\right]dp+RT\ln p. \qquad \ldots\ldots(3)$$

For gas mixtures:

Work content:

$$F_{VT}=\int_V^\infty\left[p-\frac{\Sigma_i(n_i)RT}{V}\right]dV-\Sigma_i\left(n_iRT\ln\frac{V}{n_iRT}\right)+\Sigma_i(n_if_i'). \qquad \ldots\ldots(4)$$

* Contribution from the Research Laboratory of Physical Chemistry, Massachusetts Institute of Technology, no. 460.

Free energy:

$$F_{pT} = \int_0^p \left[V - \frac{\Sigma_i(n_i)RT}{p} \right] dp + \Sigma_i(n_i RT \ln px_i) + \Sigma_i(n_i)RT + \Sigma_i(n_i f_i'). \quad \ldots\ldots(5)$$

Fugacity of gas in a mixture:

$$RT \ln f_i = \int_0^p \left[\left(\frac{\partial V}{\partial n_i} \right)_{p,t,n} - \frac{RT}{p} \right] dp + RT \ln px_i. \quad \ldots\ldots(6)$$

In the above relations

$$F_{pT} \equiv F_{VT} + pV, \quad \ldots\ldots(7)$$

$$f' = \int_{T_0}^{T} c_v^* \, dT - T \int_{T_0}^{T} c_v^* \frac{dT}{T} - RT \ln \frac{T}{T_0} + u_0' - Ts_0', \quad \ldots\ldots(8)$$

p, V, T, n, and R have their usual significance, x_i is the mole fraction of gas i in the mixture, c_v^* is the molal constant-volume heat capacity of the gas at zero pressure, and u_0' and s_0' are the molal energy and entropy constant of the gas at the standard-state temperature T_0. The subscript n to the derivative in equation (6) denotes that all of the mole numbers are to be held constant except the one occurring in the demominator.

The integrations with respect to volume and pressure are made at constant temperature and composition, and those with respect to temperature are for the pure substances and are made at zero pressure.

When the integrations have been performed, equations (1) and (2) for pure gases and equations (4) and (5) for gas mixtures are each what Gibbs called a fundamental equation. All of the properties of gases and gas mixtures can be derived from such a relation without further integration. Equations (3) and (6) for the fugacity (Lewis and Randall, 1923) of a gas are included, since this function has frequently been used by physical chemists as a convenient method of obtaining certain results.

The expression of the integrals of equation (8) in terms of elementary functions will not be discussed in the present report. Those of equations (1) or (2) can be so expressed, in principle, if we have available an equation of state for pure gases of the form $p = f(V, T, n)$ or $V = f(p, T, n)$, in which certain parameters characteristic of the gas in question occur; and those of equations (4) or (5) if we have an equation of state for gas mixtures of the form $p = f(V, T, n_1, n_2 \ldots)$ or $V = f(p, T, n_1, n_2 \ldots)$, in which parameters characteristic of the pure gases, and perhaps of the combination of gases, occur. To obtain such expressions is the problem of the equation of state.

This problem has been attacked from the standpoint of: (1) graphical methods; (2) empirical or semi-empirical equations which may have a more or less theoretical background, but are concerned mainly with the representation of the experimental compressibility data; and (3) more or less rigorous theoretical treatments based on statistical mechanics and quantum mechanics.

§ 2. GRAPHICAL METHODS

For those regions in which an equation of state gives a sufficiently accurate representation of the compressibility of a gas or gas mixture the analytic method is, in our opinion, superior to a graphical method. Graphical methods are used for the high-pressure gas region where equations of state fail to represent the behaviour of a gas and for those substances for which only a small body of data, as the critical temperature and pressure, exist. In the latter case only approximate results are possible, but in engineering and some scientific work such a solution is satisfactory.

Two graphical methods have been used. Each consists in plotting for constant temperature some function of p, V/n, T against the pressure or density. The function is so chosen that it varies slowly with the pressure or density, and hence the determination of the desired quantity, often the area under the curve, may be made with considerable accuracy.

In the first method, the relations used are

$$p = \frac{nRT}{V}\left(1 + \frac{B_v}{V/n}\right), \qquad \ldots\ldots(9)$$

$$\frac{V}{n} = \frac{RT}{p} + B_p. \qquad \ldots\ldots(10)$$

The quantities

$$B_v = \frac{V}{n}\left(\frac{pV}{nRT} - 1\right), \qquad \ldots\ldots(11)$$

$$B_p = \frac{V}{n} - \frac{RT}{p} \qquad \ldots\ldots(12)$$

are computed from the experimental data for the isotherms desired, and plots of B_v and B_p against V/n (or n/V) and p, respectively, prepared. These plots, together with equations (9) and (10), permit computation of the pressure or volume corresponding to a given volume or pressure at any temperature at which compressibility data exist, and also permit evaluation of the integrals of equations (1) to (3) as

$$\int_V^\infty \left[p - \frac{nRT}{V}\right] dV = nRT \int_0^{n/V} B_v d\left(\frac{n}{V}\right), \qquad \ldots\ldots(13)$$

$$\int_0^p \left[V - \frac{nRT}{p}\right] dp = n \int_0^p B_p \, dp. \qquad \ldots\ldots(14)$$

For computation of the energy, entropy, heat capacity and other thermodynamic properties of a pure gas, various derivatives of the type $[\partial B_v/\partial(1/T)]_{V/n}$, $[\partial(B_p/T)/\partial(1/T)]_p$ and higher temperature derivatives must be evaluated and plotted against the density and pressure for the calculation of the values of the integrals involved.

As the pressure approaches zero at constant temperature, the quantities B_p and B_v should approach the same numerical value, that of the second virial coefficient, a quantity that has been the subject of much theoretical investigation, and is

considered in some detail in later sections of this report. It should be noted that the limit of B_v (or B_p) as the pressure approaches zero is not zero, as has been assumed by several investigators using the graphical method. Both theory and the experimental data give the result that in general the value of B_v at zero pressure is finite and not zero. At the Boyle temperature and, according to theory, at infinite temperature this limit is zero.

For gas mixtures the method has been used only for computing the fugacity of a gas in a mixture by equation (6). From the experimental compressibility data on mixtures the quantity

$$B_{pi} = \left(\frac{\partial V}{\partial n_i}\right)_{T,\,p,\,n} - \frac{RT}{p} \qquad \ldots\ldots(15)$$

is computed. The fugacity of gas i in the mixture is given as

$$RT \ln f_i = \int_0^p B_{pi}\, dp + RT \ln px_i. \qquad \ldots\ldots(16)$$

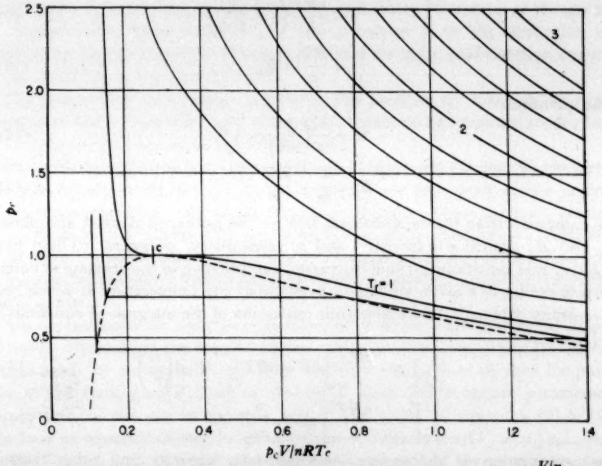

Figure 1. Plot of the reduced pressure p_r against the pseudo-reduced volume $\frac{V/n}{RT_c/p_c}$ for the reduced isotherms T_r from 1 to 3 in steps of 0·2. The critical point is marked c. The two-phase region is beneath the dotted curve. Two isotherms below the critical are sketched.

Equation (16) is not particularly useful as a general thermodynamic method nor as a method of presentation of the thermodynamic properties of gas mixtures because of the small field that can be covered by tables of fugacities of gases in mixtures. It has been used to give an indication of how well the Lewis and

Randall (1923) fugacity rule holds. This rule states that the fugacity of a gas in a mixture is equal to the product of the fugacity of the pure gas at the temperature and total pressure of the mixture into the mole fraction of the gas in the mixture. The advantage of the rule is that it permits the computation of certain thermodynamic properties of gas mixtures, such as the variation of the equilibrium constant with pressure, from data on the pure gases alone. Thus in these respects the rule plus the graphical method for pure gases is equivalent to an equation of state for mixtures.

The methods outlined above have been used by several investigators for the computation of the thermodynamic properties of pure gases and gas mixtures

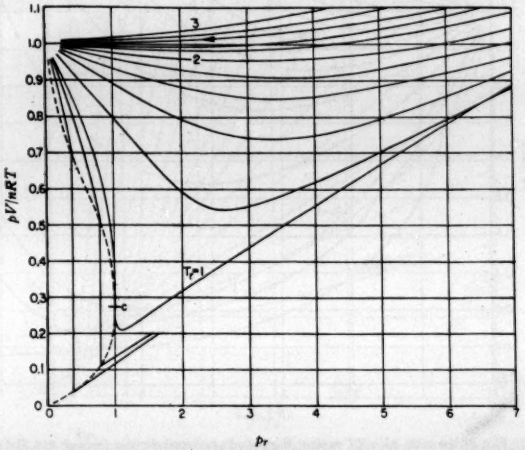

Figure 2. Plot of the ratio pV/nRT against the reduced pressure p_r for the reduced isotherms T_r from 1 to 3 in steps of 0·2. The critical point is marked c and the Boyle temperature denoted by an arrow. The two-phase region is to the left of the dotted curve. Two isotherms below the critical are sketched.

(Gibson and Sosnick, 1927; Randall and Sosnick, 1928; Merz and Whittaker, 1928; Deming and Shupe, 1931 a, 1931 b and 1932; Newton, 1935; Newton and Dodge, 1935). (See also Sage *et al.*, 1937 and 1939.)

A second graphical method makes use of the idea of a reduced equation of state (Cope, Lewis and Weber, 1931; Brown, Souders and Smith, 1932; Lewis and Luke, 1933; Lewis, 1936). It is limited to pure gases, although an extension for certain limited purposes to gas mixtures has been suggested. In figures 1 to 4 are presented certain plots, derived from compressibility data, in terms of reduced quantities. Here $p_r = p/p_c$ and $T_r = T/T_c$, where p_c and T_c are the critical

pressure and temperature (absolute) of a gas. Instead of defining a reduced volume as $V_r = V/V_c$, we define a pseudo-reduced volume, $V_r' = \frac{V/n}{RT_c/p_c}$. The reason for this is that the critical pressure and temperature are always known more accurately than the critical volume, and for many gases p_c and T_c are known whereas V_c is not. For twenty-five non-polar gases the average value of p_cV_c/nRT_c is 0·272 and of nRT_c/p_cV_c is 3·68. The maximum deviation from the value 3·68 is less than 10 per cent for non-polar gases except for hydrogen and helium.

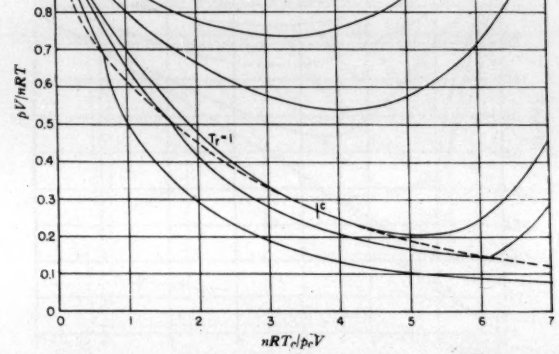

Figure 3. Plot of the ratio pV/nRT against the pseudo-reduced density $\frac{n/V}{p_c/RT_c}$ for the reduced isotherms T_r from 1 to 3 in steps of 0·2. The critical point is marked *c* and the Boyle temperature denoted by an arrow. The two-phase region is beneath the dotted curve. Two isotherms below the critical are sketched.

Van der Waals's equation gives 2·67 for this ratio, and Dieterici's equation gives 3·69.

The curves in the region from $T_r = 1$ to $T_r = 3$ in figures 1 to 3, and all those in figure 4, are drawn from the average values of the measurements for the seven hydrocarbons: methane, ethane, ethylene, propane, n-butane, isopentane, and n-heptane. They were prepared from curves drawn by Dr. Gouq-Jen Su. The curve bounding the two-phase region and the two curves below the critical were sketched from the data on steam (Keyes, Smith and Gerry, 1934 and 1936), some adjustment having been made in the neighbourhood of the critical point,

since $nRT_c/p_cV_c = 4{\cdot}30$ for this polar substance. It is interesting to note that the curves from $T_r = 1$ to $T_r = 3$ represent the compressibility of the seven hydrocarbons with an average deviation of 1 per cent, and also the compressibility of nitrogen, carbon dioxide, and steam is represented on the whole about as well by the same curves. The same curves express the compressibility of helium, hydrogen and neon approximately if we use fictitious critical temperatures and pressures for these substances (Newton, 1935).

Figure 1 is the usual plot of p against V along isotherms, but in terms of reduced quantities. Figures 2 and 3 are plots of pV/nRT against p_r and $1/V_r'$, respectively,

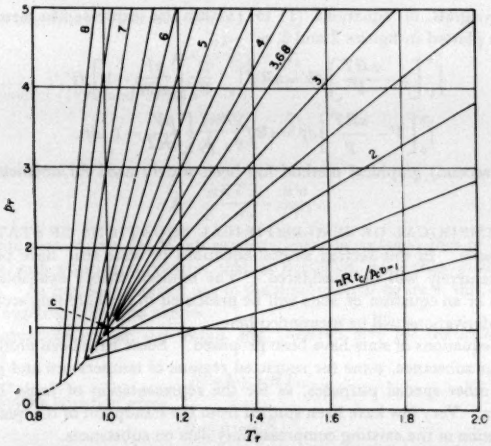

Figure 4. Plot of the reduced pressure p_r against the reduced temperature T_r for the pseudo-reduced densities $\frac{n/V}{p_c/RT_c}$ from 1 to 8 in steps of 0·5. The critical point is marked c and the pseudo-reduced density 3·68 is the critical one. The dotted curve is the vapour-pressure curve.

along reduced isotherms. In each the tangent to the curve at zero pressure changes from negative to positive at the Boyle temperature. For the curves of figure 2, where reduced pressure is the abscissa, the curvature at zero pressure also changes sign at some temperature between the critical and Boyle temperature; but for those of figure 3, where the pseudo-reduced density is the abscissa, the curvature is always positive except, perhaps, at reduced temperatures very much less than unity. It will be noted that the condensation curves in figure 3 are hyperbolae. Figure 4 is a plot of p_r against T_r along lines of constant density, that is, isometric lines. All of the early investigators were impressed by the close approach to

linearity of these curves (Ramsay and Young, 1887 a and b; Barus, 1890; Amagat, 1893). Below about the critical density the curvature of the isometrics is negative, and from about the critical density to about twice the critical density it is positive. At still greater densities the curvature is negative again. The change in sign of the curvature of the isometrics at about the critical density was noticed by Young (1895). For polar substances it seems to occur at lower reduced densities than for non-polar substances. The second change in sign of curvature at about twice the critical density has not been confirmed for all substances. Thus Amagat (1893) found positive curvature for the isometrics of methyl alcohol in the liquid region.

The integrals of equations (1) to (3) can be expressed in terms of the quantities plotted in figures 2 and 3 as

$$\int_V^\infty \left[p - \frac{nRT}{V}\right] dV = nRT \int_{V_r'}^\infty \frac{1}{V_r'} \left[\frac{pV}{nRT} - 1\right] dV_r', \quad \ldots\ldots(17)$$

$$\int_0^p \left[V - \frac{nRT}{p}\right] dp = nRT \int_0^{p_r} \frac{1}{p_r} \left[\frac{pV}{nRT} - 1\right] dp_r. \quad \ldots\ldots(18)$$

This (second) graphical method has been widely used for engineering calculations.

§ 3. EMPIRICAL OR SEMI-EMPIRICAL EQUATIONS OF STATE

Pure gases. In this section several equations of state that have been used rather extensively will be considered. The modern theory available for the derivation of an equation of state will be presented in the following section, and hence no derivations will be mentioned here.

Many equations of state have been proposed. Some have been produced for a particular substance, some for restricted regions of temperature and pressure, some for other special purposes, as for the representation of Joule-Thomson coefficients. Very few have been studied from the standpoint of the quantitative representation of the existing compressibility data on substances.

The empirical relations

$$p\frac{V}{n} = A\left(1 + \frac{nB_v}{V} + \frac{n^2C_v}{V^2} + \frac{n^4D_v}{V^4} + \ldots\right), \quad \ldots\ldots(19)$$

$$p\frac{V}{n} = A + B_p p + C_p p^2 + D_p p^4 + \ldots, \quad \ldots\ldots(20)$$

where $$A \equiv A_0 \frac{T}{T_0},$$

and the coefficients B_v, C_v, ..., B_p, C_p, ... are functions of the temperature, have been used by Kamerlingh Onnes (1901 and 1912) and the investigators at the Physikalisch-Technische Reichsanstalt, respectively (see Otto, 1929, where virial coefficients for various gases are given). Equation (19) is the so-called virial expansion for $p\frac{V}{n}$, and B_v, C_v, ... are the second, third, ... virial coefficients.

The practical equations (19) and (20) used for fitting data omit the odd powers of density and pressure except the first; theoretical considerations give all of the powers. The curves of figures 2 and 3 indicate that B_T and B_p change sign at the same temperature, the Boyle temperature; C_p changes sign between the critical and Boyle temperatures, whereas C_v is always positive, at least to temperatures well below the critical. An equation with V and T as the independent variables will give a better fit of the experimental data with the same number of terms than one with p and T as the independent variables, although the latter are preferable from an experimental point of view.

The three classical equations of state are:—

van der Waals (1899 and 1900):

$$p = \frac{nRT}{V-nb} - \frac{n^2a}{V^2}, \qquad \ldots\ldots(21)$$

Dieterici (1899 and 1901):*

$$p = \frac{nRT}{V-nb} e^{-\frac{na}{VRT}}, \qquad \ldots\ldots(22)$$

Berthelot (1907):

$$p = \frac{nRT}{V-nb} - \frac{n^2a}{TV^2}. \qquad \ldots\ldots(23)$$

The equation of van der Waals predicts linear isometrics; that of Dieterici, isometrics with positive curvature; that of Berthelot, isometrics with negative curvature. As has been pointed out in the discussion of figure 4, this curvature is always negative in the gas region.

The equations of van der Waals and Dieterici give for the second virial coefficient

$$B_T = b - a/RT, \qquad \ldots\ldots(24)$$

and Berthelot's equation gives

$$B_T = b - a/RT^2. \qquad \ldots\ldots(25)$$

Thus equation (24) predicts that a plot of B_T against $1/T$ will be a straight line, and equation (25) that a plot of B_T against $1/T^2$ will be a straight line. Actually (Keyes, 1929), when B_T is plotted against $1/T$, the curve, except for highly polar substances, is almost straight for a considerable temperature region, dropping below a straight line by progressively greater amounts as $1/T$ increases.† As $1/T$ decreases, the curve goes through a maximum rather suddenly and then approaches zero as $1/T$ approaches zero. The maximum in the curve has been found experimentally (Keyes, 1929) for helium (−75° c.), neon (300 to 400° c.) and hydrogen (450° c.), and it is estimated by him to be at 1900° c. for nitrogen.

All three equations predict a critical point, and a continuous isotherm of the type suggested by James Thomson (1871) during condensation. It will be noted

* See also MacDougall, 1916 and 1917.

† Keyes, *Temperature*, p. 45 (Reinhold Publishing Co., New York, 1941).

that each of these equations has three parameters, and since at the critical point we have, in addition to the original equation of state, the two other conditions $(\partial p/\partial V)_T = 0$ and $(\partial^2 p/\partial V^2)_T = 0$, it is evidently possible to express these three parameters in terms of the three critical constants p_c, v_c, T_c for any substance. These relations are:—

van der Waals:

$$b = v_c/3, \qquad a = 3p_c v_c^2, \qquad \frac{RT_c}{p_c v_c} = \frac{8}{3} = 2{\cdot}67;$$

Dieterici:

$$b = v_c/2, \qquad a = e^2 p_c v_c^2, \qquad \frac{RT_c}{p_c v_c} = \frac{e^2}{2} = 3{\cdot}69;$$

Berthelot:

$$b = v_c/3, \qquad a = 3p_c v_c^2 T_c, \qquad \frac{RT_c}{p_c v_c} = \frac{8}{3} = 2{\cdot}67;$$

Berthelot (empirical):

$$b = v_c/4, \qquad a = \frac{16}{3} p_c v_c^2 T_c, \qquad \frac{RT_c}{p_c v_c} = \frac{32}{9} = 3{\cdot}56.$$

For 25 non-polar gases the average value of $RT_c/p_c v_c$ is 3·68. When the values of R, b, and A expressed in terms of the critical constants are substituted into the equations (21) to (23) we obtain the corresponding reduced equation of state, that is, one containing only the reduced quantities p_r, v_r, T_r, and pure numbers which are not characteristic of the particular gas under consideration.

The equations of state of van der Waals, Dieterici, and Berthelot could be used to give a fairly satisfactory representation of the compressibility data on gases at low pressures over a short temperature range provided that the values of the parameters a and b were determined from the data themselves and not from the critical constants. The use of critical data for computing the parameters of van der Waals's and Berthelot's equations leads to particularly poor results at low pressures since the ratio $RT_c/p_c v_c$ is so far from the experimental value for real gases. Berthelot was interested in an equation for the correction of Chappuis' gas-thermometer measurements for the imperfection of the gas. For these low pressures he suggested the "empirical" set of relations given above for computing a and b from critical constants. These give a satisfactory value, 3·56, for $RT_c/p_c v_c$ and lead to the reduced equation

$$p_r v_r = \frac{32}{9} T_r + \left(1 - \frac{6}{T_r^2}\right)\frac{p_r}{4} \qquad \ldots\ldots(26)$$

for low (but not for high) pressures. This relation has found some use, but we believe that reduced plots of the type of figures 2 and 3 give better results; that is, the law of corresponding states is much better than any one of the three classical equations under consideration.

The Keyes equation of state has the form for a non-associating gas (Keyes, 1917)*:—

$$p=\frac{nRT}{V-n\delta}-\frac{n^2A}{(V+nl)^2}\,;\quad \delta=\beta\epsilon^{-na/V}. \qquad \ldots\ldots(27)$$

The values of the parameters of this equation of state have been determined for many gases (Keyes, 1914 and 1921 a; Keyes and Felsing, 1919 and 1920; Keyes, Smith, Joubert and Taylor, 1922; Smith and Taylor, 1922 and 1923 a; Beattie, 1924; and Keyes and Burks, 1927). The equation has been used to correlate and compute the thermodynamic properties of gases by Keyes and Kenney (1917), Keyes (1920, 1921 b, 1922, 1924 and 1927), Phillips (1921), Keyes, Townsend and Young (1922), Smith and Taylor (1923 b), Keyes and Sears (1925), Gillespie (1926), and Lurie and Gillespie (1927).

The Keyes equation predicts linear isometrics, and gives for the second virial coefficient the same relation, equation (24), as that derived from van der Waals's equation. Since there are five constants and three conditions at the critical point, a Keyes reduced equation of the type of equation (26) cannot be derived. This is perhaps fortunate since the values of the constants of the Keyes equation have necessarily been determined directly from the compressibility data themselves, which is certainly the proper method to obtain the best representation of the experimental results by any semi-empirical equation of state.

The Beattie-Bridgeman equation of state (Beattie and Bridgeman, 1927, 1928 a, 1928 b; Beattie, 1930 b; Scatchard, 1930) is

$$\left.\begin{aligned} p&=\frac{nRT(1-\epsilon)}{V^2}[V+nB]-\frac{n^2A}{V^2}\,;\quad B=B_0\left(1-\frac{nb}{V}\right),\\ &\qquad\qquad A=A_0\left(1-\frac{na}{V}\right),\\ &\qquad\qquad \epsilon=\frac{nc}{VT^3}. \end{aligned}\right\} \qquad \ldots\ldots(28)$$

This equation has been applied to most of the existing body of compressibility data on gases (see above references and also Gillespie (1929 a), Beattie and Lawrence (1930), Beattie and Bridgeman (1930), Beattie, Hadlock and Poffenberger (1935), Smith, Beattie and Kay (1937), Beattie, Kay and Kaminsky (1937), Beattie, Simard and Su (1939), and Lashakov (1939)). The values of the parameters for several gases are given in table 1. The equation predicts isometrics with negative curvature, as is evident when it is written in the form

$$\left.\begin{aligned} p&=T\Psi(V/n)-\Phi(V/n)-\frac{\Gamma(V/n)}{T^2}\,;\quad \Psi(V/n)=\frac{nR}{V^2}[V+nB],\\ &\qquad\qquad \Phi(V/n)=\frac{n^2A}{V^2},\\ &\qquad\qquad \Gamma(V/n)=\frac{nc}{V}\Psi(V/n). \end{aligned}\right\} \qquad \ldots\ldots(29)$$

* The form which this equation takes for an associating gas is discussed by Keyes and Taylor (1927), Beattie (1927), and Bridgeman (1927).

Table 1. Values of the constants of the Beattie-Bridgeman equation of state for several gases

$$p = [RT(1-\epsilon)/V^2][V+B] - A/V^2,$$

$$A = A_0(1-a/V), \qquad B = B_0(1-b/V), \qquad \epsilon = c/VT^3$$

Units: normal atmospheres, litres per mole, °K. (T° K. $= t$° C. + 273·13)

$$R = 0{\cdot}08206$$

Gas	A_0	a	B_0	b	$10^{-4} \times c$	Mol. wt.
He	0·0216	0·05984	0·01400	0·0	0·0040	4·00
Ne	0·2125	0·02196	0·02060	0·0	0·101	20·2
A	1·2907	0·02328	0·03931	0·0	5·99	39·91
H_2	0·1975	−0·00506	0·02096	−0·04359	0·0504	2·0154
N_2	1·3445	+0·02617	0·05046	−0·00691	4·20	28·016
O_2	1·4911	0·02562	0·04624	+0·004208	4·80	32
Air	1·3012	0·01931	0·04611	−0·01101	4·34	28·964
CO_2	5·0065	0·07132	0·10476	+0·07235	66·00	44·000
NH_3	2·3930	0·17031	0·03415	+0·19112	476·87	17·0311
CH_4	2·2769	0·01855	0·05587	−0·01587	12·83	16·0308
C_2H_6	5·8800	0·05861	0·09400	+0·01915	90·00	30·0462
C_2H_4	6·1520	0·04964	0·12156	+0·03597	22·68	28·0308
C_3H_8	11·9200	0·07321	0·18100	+0·04293	120·00	44·0616
n-C_4H_{10}	17·7940	0·12161	0·24620	+0·09423	350·00	58·077
iso-C_4H_8	16·9600	0·10860	0·24200	+0·08750	250·00	56·0616
n-C_7H_{16}	54·520	0·20066	0·70816	+0·19179	400·00	100·1248
CH_3OH	33·309	0·09246	0·60362	+0·09929	32·031	32·031
$(C_2H_5)_2O$	31·278	0·12456	0·45446	+0·11954	33·33	74·077

The curvature predicted by the equation increases with density. Hence when fitting the equation to the compressibility of polar gases for which the change in curvature of the isometrics occurs at or below the critical density, a compromise must be made in the value of c selected to represent the curvature of all of the isometrics up to the critical. This compromise sometimes causes the second virial coefficient computed from the equation to be in error by an appreciable amount. For non-polar gases for which the change in curvature may occur considerably above the critical density, a single value of c which represents quite satisfactorily the curvature of all of the isometrics up to the critical may be selected.

The virial form of equation (28) is

$$p\frac{V}{n}=RT\left(1+\frac{nB_v}{V}+\frac{n^2C_v}{V^2}+\frac{n^3D_v}{V^3}\right);\quad \left.\begin{aligned} B_v&=B_0-\frac{A_0}{RT}-\frac{c}{T^3},\\ C_v&=-B_0b+\frac{A_0a}{RT}-\frac{B_0c}{T^3},\\ D_v&=\frac{B_0bc}{T^3}. \end{aligned}\right\} \quad \ldots\ldots(30)$$

Thus the equation does not predict a maximum in the second virial coefficient B_v. Values of B_0, A_0, and c obtained by applying the equation of state (28) to all of the compressibility data for a substance permit the computation of second virial coefficients in fair agreement with those obtained by the virial expansion equation (19) for non-polar gases from low temperatures to those well above the Boyle temperature.

The limitations of the equation of state (28) may be stated thus: it gives a satisfactory representation of the compressibility of non-polar gases and a fair representation for most polar gases from low densities to about the critical density, and from low temperatures to temperatures well above the Boyle temperature.

The classical method of obtaining a reduced equation of state from equation (28) cannot be applied, but Keyes (1938) has shown that an approximate reduced equation of state, sufficient for many purposes, can be derived. For low pressures equation (30) can be written in either of two equivalent forms:

$$pV/n=RT(1+nB_v/V), \quad \ldots\ldots(31)$$

$$=RT+B_vp. \quad \ldots\ldots(32)$$

From equation (30) we may write

$$B_v\frac{p_c}{RT_c}=\frac{B_0}{R}\frac{p_c}{T_c}-\frac{A_0}{R^2T}\frac{p_c}{T_c}-\frac{c}{RT^3}\frac{p_c}{T_c}, \quad \ldots\ldots(33)$$

$$=\left(\frac{B_0}{R}\frac{p_c}{T_c}\right)-\left(\frac{A_0p_c}{R^2T_c^2}\right)\frac{1}{T_r}-\left(\frac{cp_c}{RT_c^4}\right)\frac{1}{T_r^3}. \quad \ldots\ldots(34)$$

Keyes found that for twelve non-polar gases the three quantities in the parentheses were each fairly constant, independent of the gas, and for two polar gases they were each constant again. Thus for any reduced temperature $T_r(=T/T_c)$ we may write:

For non-polar gases:

$$B_v(T_r) = \frac{RT_c}{p_c}\left[0{\cdot}1620 - \frac{0{\cdot}4253}{T_r} - \frac{0{\cdot}0667}{T_r^3}\right]. \qquad \ldots\ldots(35)$$

For polar gases:

$$B_v(T_r) = \frac{RT_c}{p_c}\left[0{\cdot}1384 - \frac{0{\cdot}3140}{T_r} - \frac{0{\cdot}1438}{T_r^3}\right]. \qquad \ldots\ldots(36)$$

Equations (35) and (36) may be considered as universal relations for B_v as functions of temperature, $B_v(T)$, and of the parameters p_c and T_c, since $T_r = T/T_c$. Equations (31) and (32), together with equations (35) and (36), have been used by Keyes for several thermodynamic calculations with surprisingly good results. These relations give much better results than the approximate reduced equation of Berthelot, equation (26), and are recommended for use in preference to other methods for approximate calculations at low pressure when only critical data are available.

The Beattie-Bridgeman equation of state has been used to compute and correlate the general thermodynamic properties of gases (Beattie and Bridgeman, 1928 c; Beattie, 1928 c, 1929 a, 1930 c and 1930 d; Bridgeman, 1929).

Recently Benedict, Webb and Rubin (1940) have published an equation of state for hydrocarbons:

$$\left.\begin{aligned} \frac{p}{d} &= T\Psi(d) - \Phi(d) - \frac{\Gamma(d)}{T^2}\,; \quad \Psi(d) = R[1 + B_0 d + bd^2], \\ \Phi(d) &= A_0 d + ad^2 - a\alpha d^5, \\ \Gamma(d) &= C_0 d - cd^2(1+\gamma d^2)\exp(-\gamma d^2), \\ d &\equiv n/V, \end{aligned}\right\} \qquad \ldots\ldots(37)$$

which gives an accurate representation of the compressibility to about twice the critical density. The changes in curvature of the isometrics at about the critical and twice the critical densities are faithfully represented. The critical region is well represented, and vapour pressures as well as saturation liquid and gas densities may be calculated from the equation with extraordinary accuracy. This equation of state gives the same relation for the second virial coefficient as that derived from the Beattie-Bridgeman equation.

Gas mixtures. Theoretical and practical considerations indicate that the equation of state of a mixture of gases that do not react chemically is of the same form as that for the pure component gases. The problem of the equation of state of gas mixtures is one of computing the values of the parameters for the mixture from those of the pure gases with or without the introduction of any new constants.

The application of the virial equation of state, equation (19), to gas mixtures has been the subject of theoretical and experimental investigation (Verschoyle, 1926; Lennard-Jones and Cook, 1927; Kleeman, 1928; Deming and Shupe, 1929). It will be considered further in the following section of this report.

In the application of equations of state of the van der Waals type to gas

mixtures, the relation suggested by van der Waals (1899 and 1900) for computing the value of any parameter, say b_m, for a binary mixture is

$$b_m = b_1 x_1^2 + 2b_{12} x_1 x_2 + b_2 x_2^2, \qquad \ldots\ldots(38)$$

where the subscript 1 refers to gas 1, 2 to gas 2, x is the mole fraction, and b_{12} is an interaction constant, characteristic of the pair of gases composing the mixture. It is highly desirable to be able to compute the value of b_m without recourse to data on mixtures, in which case the interaction constant must be related to properties of the pure gases composing the mixture. Three methods have been suggested:—

Linear combination:

$$b_{12} = \tfrac{1}{2}(b_1 + b_2), \quad \text{whence} \quad b_m = \Sigma_i (x_i b_i). \qquad \ldots\ldots(39)$$

Square-root combination:

$$b_{12} = (b_1 b_2)^{1/2}, \quad \text{whence} \quad b_m = (\Sigma_i x_i b_i^{1/2})^2. \qquad \ldots\ldots(40)$$

Lorentz combination (1881):

$$b_{12}^{1/3} = \tfrac{1}{2}(b_1^{1/3} + b_2^{1/3}), \quad \text{whence} \quad b_m = \tfrac{1}{8}\Sigma_i \Sigma_j [x_i x_j (b_i^{1/3} + b_j^{1/3})^3]. \ldots\ldots(41)$$

The relations given for b_m in equations (39) to (41) are not restricted to binary mixtures, but may be used for a mixture of any number of gases. When mole fractions, x_i, are used in these relations, V is the volume of one average mole of mixture; when the mole numbers, n, are used, V is the volume of $\Sigma_i n_i$ moles of mixture.

Theory indicates that square-root combination is the best approximation for a_m and Lorenz combination (for billiard-ball molecules) for b_m, although, because of its much greater simplicity, linear combination has almost always been used for b_m. Many thermodynamic relations for gas mixtures have been derived and a few applied to data for the van der Waals and Berthelot equations of state (Galitzine, 1890; Berthelot, 1898; van der Waals, 1899 and 1900; v. Laar and Lorenz, 1925; Trautz *et al.*, 1926; Eucken and Bresler, 1928; v. Laar, 1928 a, 1928 b and 1929.)

Keyes (1927) used square-root combinations for A_m and linear combinations for all of the other parameters in his equation of state. This equation of state for mixtures was successful not only for representing the compressibility of gas mixtures but for computing the effect of pressure on chemical equilibria, that is, in gaseous systems in which chemical reaction takes place (Gillespie, 1926; Keyes 1927; Lurie and Gillespie, 1927; Keyes and Burks, 1928).

In the applications of the Beattie-Bridgeman equation to gas mixtures so far published (Beattie, 1928 c, 1929 a and 1929 b; Gillespie, 1929 b; Beattie and Ikehara, 1930; Gillespie and Beattie, 1930 a and 1930 b), square-root combinations for A_{om}, linear or Lorentz for B_{om}, and linear combinations for all of the other parameters were used. Unpublished computations on hydrocarbon mixtures indicate that square-root combination for A_{om} and c_m, Lorentz combination for B_{om}, and linear combination for a_m and b_m give somewhat better representation of the compressibility of mixtures.

Given an equation of state for mixtures of the form $p=f(V, T, n_1, n_2, \ldots)$ we can evaluate the integral of equation (4) in terms of elementary functions without difficulty provided the integral of $p\,dV$ can be so expressed.

§4. THEORY OF THE EQUATION OF STATE

A completely satisfactory quantitative theory of the equation of state of gases requires (*a*) the correct application of kinetic theory or statistical mechanics, and (*b*) a detailed knowledge of intermolecular forces. The classical studies of the era of van der Waals were handicapped chiefly by an insufficient understanding of the second factor, which has received adequate treatment only since the advent of the new quantum theory. In the review of the present state of the theory, we shall consider the topics (*a*) and (*b*) above in order and then compare their synthesis with the experimental evidence.

The second virial coefficient. The nature of the second virial coefficient B_v of equation (19), which represents the initial departure from ideal-gas behaviour, has been known for some time; this is to be expected, since the assumption of low densities (or large volumes) introduces convenient mathematical approximations. There are many well-known and more or less accurate ways of obtaining a theoretical relation between B_v and the intermolecular forces, so that a repetition of the argument need not be given here. Some methods (Kennard, 1938) calculate the rate of transfer of momentum across a small plane of reference in the gas; this is essentially the classical method of Lorentz and Boltzmann. Others (Fowler, 1936 a) employ the virial theorem of Clausius, while still others proceed by formal reduction of the phase integral of classical statistical mechanics (Fowler, 1936 b).

All of these methods lead to the equation

$$B_v(T)=\frac{N_0}{2V}\int_{(v)}\cdots\cdots\int(1-e^{-u/kT})\,d\tau_1 d\tau_2, \qquad \ldots\ldots(42)$$

where N_0 is Avogadro's number, k is Boltzmann's constant, $d\tau_1$ and $d\tau_2$ are the volume elements of two different molecules in configuration space, and u is the mutual potential energy of this pair of molecules as a function of their space co-ordinates. The integration is to be performed over all available phase-space; that is, over the volume of the containing vessel. The result (42) is for a pure gas, in which all the molecules are similar; it is also restricted to temperature regions in which classical statistics are sufficiently valid. Mixtures and quantum effects will be mentioned later.

As written above, B_v has the dimensions of volume, and is therefore in complete correspondence with the B_v of equation (19). In alternative forms of the virial expansion, B_v as defined in (42) may have to be multiplied by RT.

For the special case where the molecules possess spherical symmetry (as, for example, in the rare gases), so that u depends not on orientation, but only on the separation r of a pair of molecules, the equation can be simplified to

$$B_v(T)=2\pi N_0\int_0^\infty(1-e^{-u(r)/kT})\,r^2\,dr. \qquad \ldots\ldots(43)$$

The infinite limit of the integral follows from the assumption of a large volume and the fact that $u(r)$ approaches zero rapidly as r increases, becoming practically negligible at separations of a few molecular diameters. (Actually $u(r)$ must approach zero more rapidly than r^{-3} if the integral is to converge; it will be seen later that modern theories of intermolecular forces satisfy this requirement.)

The comparative simplicity of the special case of spherical symmetry has caused it to be most frequently chosen for numerical calculations, which are to be treated in a later paragraph. In making such calculations it is often convenient to integrate (43) by parts, obtaining

$$B_v(T) = -\frac{2\pi N_0}{3kT}\int_0^\infty e^{-u(r)/kT}\frac{du(r)}{dr}r^3\,dr. \qquad \ldots\ldots(44)$$

Equation (44) is valid, however, only when $u(r)$ and its first derivative are continuous.

Statistical theory of the complete equation of state. A general mathematical method for deriving the complete equation of state from the phase integral was indicated some years ago by Ursell (1927), but further developments were not immediately forthcoming. Interest was revived in 1937 with the appearance of a paper by Mayer, and since then the subject has received much attention (Mayer and Ackermann, 1937; Born, 1937; Born and Fuchs, 1938; Kahn, 1938; Kahn and Uhlenbeck, 1938 Mayer and Harrison, 1938; Harrison and Mayer, 1938; Mayer and Streeter, 1939; and Mayer and Goeppert-Mayer, 1940). Although at present numerical applications would be exceedingly difficult, the theory is sufficiently developed not only to indicate the direction in which further attempts should proceed, but also to shed some light on the phenomenon of condensation.

Several different mathematical methods have been devised to obtain the equation of state from the phase integral, but all are too long and complex for detailed reproduction here. The most complete results are obtained when classical statistics may be assumed sufficiently valid, and when, in addition, the intermolecular potential energy is assumed to be additive—that is, the total potential energy of a gas containing many molecules is merely the sum of the potential energies of all the pairs of molecules. (This statement must not be misinterpreted to mean that ternary or higher encounters are neglected. A good analogy is the many-body problem in celestial mechanics, where the total potential energy of the system is the sum of the Newtonian gravitational potential energies between all pairs of bodies.) The theory of intermolecular forces is not at present able to justify the additivity assumption completely, but it is intuitively believed that this condition is very nearly fulfilled for gases composed of chemically saturated molecules.

With these assumptions, a formal reduction of the phase integral becomes possible. It is sufficient here to recall that in classical statistics the integration over momenta may be performed at once, so that the phase integral Q for a system of N identical molecules of mass m becomes

$$Q = \left(\frac{2\pi mkT}{h^2}\right)^{3N/2}\frac{Q_r}{N!}, \qquad \ldots\ldots(45)$$

where the configuration integral Q_r is given by

$$Q_r = \int_{(v)} \cdots \int e^{-U/kT}\, d\tau_1 \ldots d\tau_i \ldots d\tau_N, \qquad \ldots\ldots(46)$$

and U is the total potential energy of the system.

When the additivity property is invoked, U becomes a sum of pair-potentials u_{ij}, and $e^{-U/kT}$ a product of terms $e^{-u_{ij}/kT}$. By the simple device of defining a quantity

$$f_{ij} = e^{-u_{ij}/kT} - 1, \qquad \ldots\ldots(47)$$

this product may be factorized into a sum of simpler products of the f_{ij}. The advantage of this procedure lies in the fact that each term in the sum may be treated separately, and that the terms can be defined in a general manner and classified according to the number of molecules whose potential energies appear in the term.

It is impossible to discuss here the details of the combinatory analysis or complex-function theory by which the final result is achieved. After classifying the terms in Q_r and applying the usual equation,

$$p = -\left(\frac{\partial A}{\partial V}\right)_T = kT\left(\frac{\partial \ln Q_r}{\partial V}\right)_T, \qquad \ldots\ldots(48)$$

it is found that for sufficiently large volumes the equation of state of the gas may be written

$$pV/nRT = 1 - \sum_{k \geq 1} \frac{k}{k+1} \beta_k (n/V)^k. \qquad \ldots\ldots(49)$$

where the β_k are defined in a general manner and are functions of temperature. Thus the experimentalists' practice of expanding pV in the inverse volume (equation (19)) is theoretically countenanced, although there is no evidence for the omission of higher odd powers, which has been found to be economical in fitting experimental results.

The general definition of β_k will not be given here. The first three terms are

$$\begin{aligned}
\beta_1 &= \frac{N_0}{V} \int\int f_{12}\, d\tau_1 d\tau_2, \\
\beta_2 &= \frac{N_0^2}{2V} \int\int\int f_{12} f_{23} f_{13}\, d\tau_1 d\tau_2 d\tau_3, \\
\beta_3 &= \frac{N_0^3}{6V} \int\int\int\int (3 f_{12} f_{23} f_{34} f_{14} + 6 f_{12} f_{23} f_{34} f_{14} f_{13} + f_{12} f_{23} f_{34} f_{13} f_{24})\, d\tau_1 d\tau_2 d\tau_3 d\tau_4,
\end{aligned} \qquad \ldots\ldots(50)$$

and the complexity increases rapidly beyond this point. Since $-k\beta_k/(k+1)$ represents the $(k+1)$th virial coefficient, it is seen that $\beta_1 = -2B_v$, giving agreement with (42).

In the earlier discussion of (42), B_v was assumed to be independent of volume because consideration had been restricted to low densities. The question might

reasonably be raised as to whether the quantities β_k may be assumed independent of volume when the latter becomes smaller. The reassurance lies in the fact that the significant theoretical quantity is not the actual volume of the container, but rather the average volume per molecule. For a description of macroscopic experiments the theory therefore legitimately allows both the number of molecules and the volume to approach infinity, while keeping their ratio finite. Under these conditions the β_k remain independent of volume until the density approaches that of the liquid phase. Thus, for the gas phase, the virial expansion, equation (49), may be considered as rigorously deduced, and the coefficients are explicitly defined, though scarcely in a form suitable for numerical applications.

The theory of condensation and the critical region. It can be rigorously shown that, if certain mathematical conditions are satified by the β_k of (49) or by functions related to them, the pressure will remain constant when the volume per molecule (or per mole) is decreased below a certain value (Kahn, 1938; Kahn and Uhlenbeck, 1938). It may appear surprising that an isotherm with two analytically different parts can be predicted from the classical phase integral without recourse to Thomson's " rule of equal areas" (Thomson, 1871), but it has been pointed out (Kahn, 1937; Kahn and Uhlenbeck, 1938) that this result can follow when the total volume and total number of molecules are allowed to approach infinity while their ratio remains finite. Accurate specification of the *physical* conditions (i.e. temperature) under which such condensation can occur is not possible without examining the detailed dependence of the β_k on the temperature. A glance at (50) shows that this problem would present some formidable mathematical difficulties, even if knowledge of the intermolecular forces were very complete. Some rough results have, however, been achieved (Mayer *et al.*, 1937, 1938, 1939, 1940). In spite of this obstacle, it is to be regarded as highly satisfying that rigorous analysis of the phase integral alone permits condensation as a mathematical possibility.

At present the theory is not sufficiently well advanced to predict the third portion of the isotherm, corresponding to the single liquid phase. It is believed that the volume dependence of the β_k at high densities must be examined in order to make this prediction (Mayer *et al.*, 1937, 1938, 1939, 1940; Kahn, 1938; Kahn and Uhlenbeck, 1938). Consequently a rigorous statistical theory of liquids still seems rather remote at present.

A formal method of procedure has been indicated (Kahn, 1938, and Kahn and Uhlenbeck, 1938) even for the case where quantum effects are important and the potential energies are not additive. In this case, condensation is possible as before, but it has not yet been possible to arrive at the complete virial equation of state. The quantum theory of B_2, however, has been completed (Uhlenbeck and Beth, 1936 and 1937; Gropper, 1937).

The rigorous theory thus far discussed has been supplemented in an attempt to discover the nature of the critical region (Mayer *et al.*, 1937 to 1940). The classical concept of the critical point is that at a certain temperature T_c there

exists a volume V_c for which $\left(\frac{\partial p}{\partial V}\right)_T$ and $\left(\frac{\partial^2 p}{\partial V^2}\right)_T$ are simultaneously zero; that for $T > T_c$, $\left(\frac{\partial p}{\partial V}\right)_T$ is everywhere negative, and that below T_c there is always a discontinuity in $\left(\frac{\partial p}{\partial V}\right)_T$ at the saturation gas volume. The rigorous treatment of the phase integral just discussed, however, indicates that condensation can set in as the result of either of two distinct and alternative mathematical conditions. One of these conditions corresponds to the classically conceived behaviour of $\left(\frac{\partial p}{\partial V}\right)_T$ at the saturation vapour volume, but the other makes $\left(\frac{\partial p}{\partial V}\right)_T$ approach zero continuously and has no classical counterpart. It has already been stated that the theory at present cannot accurately specify the actual physical conditions (temperature) corresponding to either of the aforementioned mathematical conditions. Mayer and his collaborators (1937 to 1940) believe that both types of condensation are physically realizable, basing their arguments on intuitive physical reasoning as to the temperature-dependence of the higher β_k of (49). Thus there is pictured a finite temperature interval $T_m \leqslant T \leqslant T_c$, in which $\left(\frac{\partial p}{\partial V}\right)_T$ goes continuously to zero as saturation is approached, but after which the pressure remains constant during a finite change in volume. The temperature T_c corresponds exactly to the old definition of the critical point, while T_m is the temperature below which $\left(\frac{\partial p}{\partial V}\right)_T$ has a discontinuity at the saturation vapour volume. It is also made probable that T_m is the temperature at which the interfacial tension becomes zero. The anomalous region must therefore be quite small, although the theory at present can make no accurate predictions about this point.

There are several experiments which seem to support such considerations (Callendar, 1928 ; Tapp, Steacie and Maass, 1933; Maass, 1938). It appears that density differences may persist for quite some time at temperatures above that at which the interfacial meniscus disappears, even when there is mechanical stirring. Also, estimates of critical temperatures from the shape of the isotherms are often higher than those at which the meniscus disappears, the difference sometimes being of the order of several degrees. Nevertheless, the well-known effect (Kuenen, 1907) of small amounts of impurity in the critical region causes any such appeal to experiment to be regarded with caution. In this connection, recent measurements on very pure samples of isobutene may be of interest. From the shape of the isotherms, Beattie, Ingersoll and Stockmayer (1940) took the critical temperature to be 144·73 ± 0·02° c., while entirely independent observations of the disappearance of the meniscus gave 144·6 ± 0·6° c. (Benedict, private communication), both samples of the material having come from the same source.

It therefore still seems possible that the anomalous experiments may be the result of impurity or of failure to attain true equilibrium. The theoretical answer to the problem must await the numerical evaluation of the β_k.

Simpler theories of condensation. It has already been mentioned that the various terms in Q_v are classified in Mayer's theory according to the number of molecules simultaneously interacting, and, indeed, the integrals occurring therein are customarily referred to as "cluster integrals". When the temperature is low enough for all the β_k to be positive, corresponding to the complete predominance of the attractive forces, it is possible to write equation (49) in a different way as

$$pV/kT = \sum_{j\geqslant 1} m_j, \qquad \ldots\ldots(51)$$

where m_j is the number of "clusters" containing j molecules each, so that $\Sigma j m_j = N$ (Mayer *et al.*, 1937 to 1940). For this case the "clusters", originally conceived only as a pictorial aid in analysing the terms of Q_v, assume a simple physical meaning, and the equation of state resembles that of a mixture of perfect gases, each "cluster" behaving like one single molecule of a perfect gas. The "equilibrium constants" governing the association and dissociation of these clusters are still complicated integrals resembling the β_k of (50), however, so that no greater aid in numerical calculations is thereby attained. Several investigators (e.g. Band, 1939; Frenkel, 1939; Tseng, Feng, Cheng and Band, 1940) have taken the next natural step and treated these "equilibrium constants" in the manner appropriate to chemical association and dissociation, giving them the type of temperature-dependence shown by chemical equilibrium constants. The resulting equations can be shown to lead to condensation much in the manner of the rigorous treatment.

It is claimed (Band, 1909; Tseng *et al.*, 1940) that this quasi-chemical method is more suitable for numerical calculations; thus far, however, these have been of a highly approximate nature. In particular, the surface tension of the liquid may be roughly expressed by the same theory, and the properties of the saturated vapour then calculated from the observed values of the surface tension. Agreement with experiment is rather unsatisfactory. A more refined calculation, which employs observed values of both surface-tension and saturation vapour-volume, affords a means of estimating the average number of molecules per cluster. As might be expected intuitively, the result is that the clusters remain rather small until saturation is approached, when their average size increases rapidly.

The effect of gravity in the critical region has also been considered by the same authors by the quasi-chemical method, and it is indicated that this effect alone may be sufficient to explain the experimental results without the introduction of the anomalous temperature interval postulated by Mayer and his co-workers.

Intermolecular forces. Before the statistical theories just sketched can be compared with experiment, it is necessary to review briefly the present theory of intermolecular forces. The properties of matter in general indicate that

molecules exert attractive forces on one another except at small distances of separation, when large repulsive forces set in rather suddenly. Satisfactory qualitative explanations for many properties of matter can be given if the molecules are very simply represented as hard elastic spheres of finite magnitude which attract each other until they touch.

With the advent of modern theories of atomic structure, however, the classical concept of the hard sphere lost its physical reality, and, indeed, it was not until 1927 that Heitler and London succeeded in explaining the apparent "size" of atoms and molecules. These repulsive forces depend directly on the exclusion principle, which prevents more than two electrons from having the same quantum numbers (i.e. average charge distributions), and not even two unless their spins are anti-parallel. If on the approach of two atoms a pair of electrons, one from each atom, is permitted to stay in the same "shell", the result is an attractive force. This is the ordinary homopolar or electron-pair valence bond. If, on the other hand, both approaching atoms have closed shells, the exclusion of the electrons from the same region of space results in a repulsive force. Since electronic charge distributions always show a roughly exponential falling-off with distance from the nucleus, these interatomic repulsive forces may also be expected to show a roughly exponential (and hence very sharp) dependence on the internuclear distance. This sudden and rapid increase of repulsion with decreasing distance accounts for the success of the hard-sphere model, which corresponds to an infinite repulsive force on contact of the two spheres.

Although the qualitative nature of this repulsive or "overlap" energy is thus easily understood, its quantitative estimation is always extremely laborious and usually almost impossible, because it depends so intimately on the wave functions (i.e. charge distributions), especially at fairly large distances from the atomic nucleus. Hence detailed calculations have thus far been performed only for atomic hydrogen (Heitler and London, 1927; Sugiura, 1927), for helium (Slater, 1928; Margenau, 1939a), and for neon (Bleick and Mayer, 1934), but these have been favourably compared with experiment when possible.

There has recently been suggested an approximate method of introducing the exclusion principle (the "method of orthogonalization"), which simplifies such calculations and seems to give promising results (Mamotenko and Schuchowitzky, 1938; Mamotenko, 1939). Good agreement is obtained with the more accurate calculations for atomic hydrogen and helium, and the results for neon and argon, although not reported in great detail, seem to be in encouragingly good accord with the crystal data for these substances.

The Thomas-Fermi statistical theory of the atomic electron cloud has also been applied to the problem of the repulsive energy, and seems to give quite well the repulsive forces in the RbBr crystal (Gombás, 1935; Jensen, 1936). One may expect this method to be more successful for larger atoms because of its statistical nature. No calculation of repulsive forces between diatomic or larger molecules has yet been made, though for hydrogen this would not appear to be wholly impossible.

No simple relation between the repulsive energy and any other experimentally determinable atomic properties has yet been proposed, and, indeed, this does not seem possible without too many over-simplifying assumptions as to the nature of the atomic wave-functions. Consequently the prevailing custom is to assign to the repulsive energy a simple exponential or inverse-power dependence on the separation and then determine the unspecified parameters by comparison with experiment (crystal data, gaseous diffusion and viscosity, or equation of state).

Comparatively greater progress has been possible in the theory of the attractive forces which persist at larger distances of separation, the so-called van der Waals forces. Several reviews (London, 1937; Margenau, 1939 b) of this subject have recently been given, so that the discussion here will be quite restricted.

Older theories of van der Waals forces quite naturally attempted to identify them with electrostatic forces between less-than-spherically symmetrical charge distributions (dipoles, quadripoles, etc.) in the isolated atoms of molecules. Such theories, however, were unable to account for any attractive forces between the rare-gas atoms after it became known that their ground-state charge distributions are spherically symmetrical. It was not until 1930 that London (1930 a and 1930 b) succeeded in explaining the origin of these forces, which, indeed, are the most important attractive forces between all molecules except the very polar ones such as NH_3, H_2O, or HCN.

Once again the physical principle is easily understood qualitatively. In a rare-gas atom, the electron distribution has spherical symmetry about the nucleus only when averaged over time ; instantaneously the electrons are at arbitrary positions and give a dipole moment which is zero only on the average. When two such atoms are brought near to each other, all positions of the electrons will no longer be equally probable, but those of lowest energy, corresponding to an attraction between the two instantaneous dipoles, will become more frequent. On the average, the electrons of one atom will spend more time in those regions which are near the other atom. The result of this effect is an attractive energy whose leading term varies as the inverse sixth power of the internuclear separation; the next contribution varies as r^{-8}, the next as r^{-10}, and so on.

Accurate calculation of this attractive energy again requires a detailed knowledge of the wave functions, and again it is natural that atomic hydrogen (Eisenschitz and London, 1930; Slater and Kirkwood, 1931; Pauling and Beach, 1935) and helium (Hassé, 1930 and 1931; Slater and Kirkwood, 1931; Baber and Hassé, 1937) have been most frequently studied. However, several valuable approximate relationships between this attractive energy and other atomic properties have been developed, and these permit good estimates of its magnitude for larger atoms and even for fairly complex molecules. By treating the atoms as harmonically oscillating dipoles London (1930 a and b) obtained for the leading term, at large distances, the formula

$$u(r) = -3h\nu\alpha^2/4r^6, \qquad \ldots\ldots(52)$$

where r is the internuclear separation, α is the atomic polarizability, and $h\nu$

is an energy which may be identified either with the ionization potential (London, 1930) or with the frequency which best reproduces the dispersion curve (Margenau, 1939 b). From the latter connection these so-called "London forces" are sometimes also called "dispersion forces". The relation (52) is valid only for such large distances that $r^3 \gg \alpha$.

With the harmonic oscillator model of the atom just mentioned, the wave equation can be solved exactly, and (52) follows from an expansion of the energy expression thus obtained. A slightly better relation may be derived by assuming a better atomic model (that of one-electron wave functions of unspecified form) and using the variation method (Hassé, 1930 and 1931; Slater and Kirkwood, 1931; Kirkwood, 1932; Baber and Hassé, 1937). In its simplest form, this calculation gives for the leading term

$$u(r) = -\frac{3}{2} n^{1/2} \alpha^{3/2} E_0 a_0^{3/2} r^{-6}, \qquad \ldots\ldots (53)$$

where E_0 is the atomic unit of energy (equal to the magnitude of the ground-state energy of the hydrogen atom, or $e^2/2a_0$), a_0 is the radius of the first Bohr orbit in hydrogen, and n is a pure number whose interpretation has been somewhat disputed. If all the electrons in the atom are regarded as contributing equal shares to the instantaneous polarization, n becomes the total number of electrons in one atom (Kirkwood, 1932). If, on the other hand, only the electrons in the outermost shell are considered to have any appreciable effect, n is taken as the number of electrons in the outer shell (Slater and Kirkwood, 1931). Undoubtedly the correct value of n must lie between these limits, and is probably much closer to the smaller value. Fortunately this ambiguity in n is not serious unless the atoms are quite large, in which case the result may be greatly improved by estimating the contribution of each shell separately (Hellmann, 1935; Buckingham, 1937).

Substitution of the numerical values for the universal constants in (53) gives

$$u(r) = -1{\cdot}24 \times 10^{-23} n^{1/2} \alpha^{3/2} r^{-6}, \qquad \ldots\ldots (54)$$

which yields good agreement with the more accurate calculations for helium when the experimental value of α is substituted therein.

The second term in the attractive energy, which varies as r^{-8}, has also been examined and approximate expressions for it have been developed (Buckingham, 1937; Margenau, 1939 b). The harmonic oscillator model gives the particularly simple result of Margenau (1939 b),

$$u^{(2)}(r) = -\frac{15}{2} E_0 \alpha^2 a_0^2 r^{-8}, \qquad \ldots\ldots (55)$$

where the symbols have the same meanings as in (53). Comparison of (55) with (53) indicates that often the second term is of a size comparable to the first at significant values of r, so that for accurate calculations it should not be neglected. This explains why the use of the total number of electrons for n in (53) often gives better agreement with experiment than the theoretically more preferable, but smaller, number of electrons in the outer shell alone.

For molecules which are not spherically symmetrical the use of such simple formulae as (52) or (53) may be questioned. A recent detailed calculation by Massey and Buckingham (1938 a) of the attractive energy between two H_2 molecules, however, yields the value $1{\cdot}43 \times 10^{-59}$ erg-cm^6 for the coefficient of r^{-6} after averaging over all orientations, while the simple formula (53) gives $1{\cdot}30 \times 10^{-59}$.

The complete interatomic energy as a function of internuclear distance is customarily obtained by grafting together the entirely independent calculations for the overlap energy and the dispersion energy. Although such a procedure may yield good results in practice, its theoretical justification in general is very doubtful, since in the most important region (which includes the zero and minimum of the energy curve) the assumption of large distances required for the validity of (52) or (53) is not sufficiently fulfilled. Its empirical success is very probably explained by the short range and rapid ascent of the repulsive energy. Recently this question has been considered for the particular cases of atomic hydrogen and helium; the procedure is shown to be very bad for H, but quite permissible for He (Margenau, 1939 a).

Comparisons of theory with experiment. Because of mathematical difficulties, most attempts to correlate theory with experimental equations of state have necessarily been confined to a calculation of the second virial coefficient for spherical molecules by the use of equation (43). The simplest possible molecular model is the rigid non-attracting sphere of diameter σ, which corresponds to infinite potential energy for $r \leqslant \sigma$, and zero potential energy for $r > \sigma$. Substitution in (43) gives $2\pi N_0\sigma^3/3$ for B_φ. The hopeless inadequacy of this model soon leads to the next logical step, the introduction of an attractive potential for distances $r > \sigma$. One simple assumption is to give $u(r)$ the constant value $-D$ in a region $\sigma < r \leqslant \rho$ (the so-called "box potential"), which gives

$$B_\varphi = 2\pi N_0\sigma^3/3 - (2\pi N_0/3)(\rho^3 - \sigma^3)(e^{D/kT} - 1). \qquad \ldots\ldots(56)$$

Another, and more reasonable, assumption is to let the attractive energy vary inversely as some power of r—that is, letting $u(r) = -\mu r^{-m}$ for $r > \sigma$. The sixth power is indicated theoretically, as we have seen. Integration then gives

$$B_\varphi = \frac{2\pi N_0\sigma^3}{3}\left\{1 - \sum_{n \geqslant 1} \frac{3(\epsilon/kT)^n}{n!\,(nm-3)}\right\}, \qquad \ldots\ldots(57)$$

where $\epsilon = \mu\sigma^{-m}$ (Keesom, 1912).

Either (56) or (57) can be made to fit observed values of B_φ quite well for all but the lightest non-polar gases by adjusting the three parameters σ, ρ, D or σ, μ, m. For the second case m can generally also be set at its preferred value of 6.

If D/kT or ϵ/kT in the above formulae are assumed small, the neglect of higher terms gives a result of the form

$$B_\varphi = b - a/RT, \qquad \ldots\ldots(58)$$

which is seen to be exactly the expression (24) given by the semi-empirical equations of van der Waals, Dieterici, and Keyes. The Beattie-Bridgeman expression for B_v, given by equation (30), corresponds to representing the higher terms in the series by c/T^3.

If m in (57) is set at 6, and higher terms are neglected for comparison with (58), it is found that

$$b = 2\pi N_0\sigma^3/3, \quad a = N_0 b\epsilon. \qquad \ldots\ldots(59)$$

It has already been stated that estimation of repulsive energy (or of σ for this case) by theory is very difficult. Therefore tests of (52) or (53) have very often been made by taking experimental values for b and then calculating a by (59), using the theoretical value for ϵ. Rough agreement is usually attained, which is about all that can be expected when the approximations in (59), as well as in (52) or (53), are recalled. It is also well to remember that experimental values of a and b are in general chosen not for the best fit of B_v alone, but to give a good average representation of a large body of data, if not to reproduce the critical temperature and pressure.

Satisfactory as this hard-sphere model is for most gases, its shortcomings are apparent when the data for helium and neon are examined, for these substances exhibit a maximum in B_v, a property which is clearly excluded by (57). An improved model which yields to analytical treatment was proposed in 1924 by Lennard-Jones, who represented the repulsive energy by an inverse-power dependence on the separation. The complete potential energy was written as

$$u(r) = \lambda r^{-s} - \mu r^{-t}, \quad t < s. \qquad \ldots\ldots(60)$$

Integration of (43) with this expression for $u(r)$ is straightforward, and gives the following result:—

$$B_v = \frac{2}{3}\pi N_0(\lambda/\mu)^{3/(s-t)}F(y), \qquad \ldots\ldots(61)$$

where

$$F(y) = y^{3/(s-t)}\left\{\Gamma\left(\frac{s-3}{s}\right) - \sum_{n=1}^{\infty} c_n y^n\right\}, \qquad \ldots\ldots(62)$$

and the parameter y is defined by

$$y = (\mu/kT)\,(kT/\lambda)^{t/s}, \qquad \ldots\ldots(63)$$

the coefficients c_n being given by

$$c_n = 3\Gamma\left(\frac{nt-3}{s}\right)\Big/ n!\, s. \qquad \ldots\ldots(64)$$

This expression possesses all the required qualitative properties of the experimental B_v values, and by proper adjustment of the parameters it can be made to fit almost all data within experimental uncertainty, even when t is set at the theoretically preferred value of 6. The accuracy of representation is about equal with any value of s between 9 and 14, so that $s = 12$ is almost always assumed for reasons of mathematical simplicity, leaving only the parameters λ and μ to be

adjusted. The values of μ thus obtained show rather good agreement with (52) or (53) for a number of gases. Several excellent summaries of the Lennard-Jones treatment can be found (Lennard-Jones, 1931; Fowler, 1936c; Fowler and Guggenheim, 1939).

Although equations (61) to (64) are quite complicated, approximate equations can be developed which greatly facilitate the labour involved. For the special case $t=6$ and $s=12$, which is sufficient for most purposes, a useful equation is

$$B_0/\beta = 0{\cdot}71875 - 2{\cdot}131899\,\Theta/T - 1{\cdot}120795(\Theta/T)^2, \quad \ldots\ldots(65)$$

where

$$\beta = \frac{2}{3}\pi N_0(\lambda/\mu)^{1/2}$$

and

$$\Theta = \mu^2/4k\lambda.$$

This equation is sufficiently close to the true expression for comparison with experiment between $\Theta/T=0{\cdot}25$ and $\Theta/T=0{\cdot}9$, which covers a temperature range from slightly below the critical temperature to well above the Boyle temperature, and thus spans the experimentally accessible region for all but the lightest gases.

A second equation, almost as accurate as (65), and covering the same temperature interval, is particularly suited for very rapid estimates of Θ and β, and hence of λ and μ. The equation is

$$B_0/\beta = (\Theta/T)^{1/4}[1{\cdot}064 - 3{\cdot}602\,\Theta/T], \quad \ldots\ldots(66)$$

and fairly good values of Θ and β can be very quickly derived merely by plotting the experimental values of $B_0T^{1/4}$ against $1/T$ and identifying the best straight line through the points with (66).

Even the Lennard-Jones expression for the intermolecular energy is not quite accurate enough to give complete agreement with the very best experimental data, especially when both gas and crystal data for the same substance are considered. The defect may be partly due to the omission of the term in r^{-8} in the attraction, but the chief fault is probably in the repulsive energy, for which exponential rather than inverse-power dependence on r is theoretically indicated. By the device of dividing the temperature range of observation into small regions and obtaining the best value of λ for each region, while keeping the same value of μ throughout, the constants for a potential energy

$$u(r) = Ae^{-br} - \mu r^{-6} \quad \ldots\ldots(67)$$

have been obtained for helium, neon, and argon, and the expressions so obtained also give good values for the lattice energies and lattice constants of the crystals of neon and argon (Buckingham, 1938). It is, unfortunately, not possible to obtain an analytical expression for B_0 by the direct use of (67).

The only completely *a priori* calculation of B_0 has been for helium (Kirkwood and Keyes, 1931). An expression of the form of (67) was employed, the parameters being taken entirely from quantum-mechanical calculations by Slater (1928) and Slater and Kirkwood (1931), and values of B_0 were obtained by numerical integration which gave excellent agreement with experiment for

temperatures so high that quantum corrections were small. One disquieting feature, however, is that the introduction of a theoretically calculated r^{-8} term in the attraction considerably impairs the agreement, so that some fortuitous cancellation of errors must have occurred.

Non-polar molecules which are not spherically symmetrical, such as H_2, N_2, and C_2H_6, generally have B_v-values which can be fitted by the Lennard-Jones expression over fairly wide ranges of temperature, but highly polar gases like H_2O and NH_3 show a much greater variation of B_v with temperature, due mainly to the great dependence on orientation of the forces between permanent dipoles. Although a complete and correct comparison of theory and experiment has not yet been published, the proper inclusion of both electrostatic and dispersion forces gives an expression for B_v which gives a good fit of the data with entirely reasonable values of the parameters (Stockmayer, 1940). Because of the magnitude of the attractive forces, a hard-sphere repulsive force is entirely satisfactory in this case.

Although space does not permit a full account, it may be mentioned that the assumption of classical statistics is not valid for the lightest gases (He, H_2, Ne) at very low temperatures. The correct quantum theory of B_v has been given (Uhlenbeck and Beth, 1936 and 1937; Gropper, 1937) and has several times been compared with experiment for helium (Massey and Buckingham, 1938 b; de Boer and Michels, 1938 and 1939 a). An interesting point is that at low temperatures the B_v-values of deuterium differ appreciably from those of hydrogen, since the masses play an important part in the quantal formulae (Schäfer, 1937 a and 1937 b). Quantum effects in the polar gases have also been considered, and these are found to be quite small, though not entirely negligible at ordinary temperatures (Kirkwood, 1933).

Calculation of higher virial coefficients has been very rare because of the difficulty of integration. The third and fourth virial coefficients for rigid non-attracting spheres were long ago calculated by Boltzmann and others (see Happel, 1906, where references to earlier work are given), the results being

$$C_v = \frac{5}{8}b^2, \quad D_v = 0{\cdot}2869b^3, \qquad \ldots\ldots(68)$$

where $b = 2\pi N_0\sigma^3/3$. It has recently been suggested that at high temperatures these values may serve as useful approximations in the absence of better knowledge (Hirschfelder and Roseveare, 1931). By using the Lennard-Jones result for B_v and taking b as an additional arbitrary constant, fair agreement with high-pressure Joule-Thomson data on argon is secured, but several trends in the data are not reproduced. For one thing, such an equation cannot predict the change in sign of curvature of the isometrics near the critical density, which requires temperature dependence of the higher virial coefficients.

The third virial coefficient for the Lennard-Jones potential, equation (60), with $s = 12$ and $t = 6$, has recently been evaluated numerically for temperatures

above the critical (de Boer and Michels, 1939b). Numbers were not reported, but the equation

$$C/\beta^2 = 0{\cdot}268 + 0{\cdot}086\ \Theta/T + 0{\cdot}488(\Theta/T)^2 \qquad \ldots\ldots(69)$$

reproduces the given curve well, β and Θ having the same meaning as in (65) and (66). This result agrees quite well with experimental values for argon and nitrogen. Mayer and his co-workers (1937 to 1940) have also made some calculations of higher virial coefficients, but thus far only very rough results have been published. More work on the higher virial coefficients is much to be desired.

Corresponding states. The recent return to favour of the law of corresponding states for practical purposes, as set forth in an earlier part of this report, has also revived some theoretical interest in it. This law may be shown to be rigorously true for a gas obeying classical statistics when, and only when, the intermolecular potential energy has the same analytic form for all gases and permits just two independent constants for each gas. An expression satisfying this requirement would be the Lennard-Jones formula, equation (60), with the same values of s and t for all gases. When the approximate nature of such two-parameter expressions for the intermolecular energy is recalled, the empirical success of corresponding states is truly striking. The failure of helium and hydrogen to correspond very well to the heavier gases is easily understood, since at their critical points the quantum effect is large, and a third parameter, the mass, has an important effect.

An interesting approximate theory of dense gases by Lennard-Jones and Devonshire (1937, 1938, 1939) has succeeded in predicting quite accurately one particular corresponding-states relation, namely, that T_r (the reduced temperature) should be about 2·6 at the Boyle temperature (that at which $B_v = 0$) for gases obeying the Lennard-Jones potential with $s = 12$, $t = 6$ (see also Lennard-Jones, 1937). This is in excellent agreement with observed values, the mean for Ne, A, O_2, CO, CH_4, and N_2 being 2·63. This theory does not proceed rigorously from the phase integral, but begins with the physical argument that in a dense gas a molecule will almost always be imprisoned by its nearest neighbours and can rarely migrate very rapidly. By employing equation (60) with $s = 12$, $t = 6$, to calculate the average force on a molecule in such a cell formed by its neighbours, an approximate value for the phase integral can be found. Although critical temperatures are predicted with success, the model is too crude to give a satisfactory dependence on volume; in particular, it makes $p_c V_c/nRT_c$ equal to about 0·7, in contrast to the observed values near 0·27. In this respect it is therefore worse than van der Waals's equation.

Such a model might be expected to apply equally well, or even better, to condensed phases, and, indeed, it has been applied with success to the calculation of boiling points, and even of melting points. It is also interesting to note that it confirms the Trouton and Hildebrand rules for the entropy of vaporization.

The theory gives relations between the critical constants and the parameters

of the Lennard-Jones potential. In fact, the aforementioned prediction that $T_c = 2{\cdot}6$ at the Boyle temperature was merely a restatement of the relation $\Theta/T_c = 3/4$, where Θ has the same meaning as in (65). The volume predictions are bad, as was already mentioned. Examination of experimental data has led de Boer and Michels (1938) to propose the relations

$$V_c/n \sim 1{\cdot}5\beta, \quad T_c \sim 1{\cdot}26\Theta. \qquad \ldots\ldots(70)$$

With the help of these relations the intermolecular energy may be roughly estimated from the critical data. Conversely, the use of (70) together with (65) or (66) and with (69) should yield a fairly satisfactory reduced equation of state for densities up to about half the critical. Such an equation has been shown to give a quite satisfactory reduced Joule-Thomson inversion curve for low densities (Corner, 1939).

The approximate truth of the law of corresponding states suggests that there may be rough relationships between several of the Beattie-Bridgeman constants of equation (28). The assumption that the Beattie-Bridgeman equation gives B_v values in accord with the Lennard-Jones expression for $s = 12$, $t = 6$ between the critical and Boyle temperatures leads to the relation (Hirschfelder and Roseveare, 1939)

$$R^3 B_0^2 c/A_0^3 = 0{\cdot}0236. \qquad \ldots\ldots(71)$$

Manipulation of the correlative empirical equations (34) and (35) due to Keyes yields the value 0·0227, which is surprisingly good agreement.

Gaseous mixtures. The success of the semi-empirical combining rules (Lorentz, 1881; van der Waals, 1899 and 1900; Keyes, 1927) proposed for the Beattie-Bridgeman and other explicit equations of state suggests that they have some theoretical basis.

For the virial equation of state (19), it might be conjectured, and has been shown rigorously for a two-component system (Mayer, 1939), that the nth virial coefficient is a polynomial of nth order in the mole fractions, the coefficients being functions of temperature. The second virial coefficient, to which discussion is here restricted, is thus given for a two-component system by

$$B_v\,(\text{mixture}) = x_1^2 B_{11} + 2x_1 x_2 B_{12} + x_2^2 B_{22}, \qquad \ldots\ldots(72)$$

where the x_i are the mole-fractions and (for spherically symmetrical molecules)

$$B_{ij} = 2\pi N_0 \int_0^\infty (1 - e^{-u_{ij}(r)/kT}) r^2\, dr. \qquad \ldots\ldots(73)$$

In (73) u_{ij} is the mutual potential energy between a molecule of species i and one of species j. The theoretical problem thus reduces to that of determining u_{12} from u_{11} and u_{22} when these are known. If this can be done without recourse to any other knowledge, the principle of employing combining rules such as those of equations (38) to (41) (although perhaps not the particular rules thus far suggested) may be considered as justified theoretically.

The theory of intermolecular forces is not sufficiently developed to give very complete answers to this question. There are at present no published calculations

of the repulsive force between unlike molecules, although it should not be too difficult by the simplified "method of orthogonalization" (Mamotenko and Schuchowitzky, 1938; Mamotenko, 1939). The hard-sphere model requires additive collision diameters, and thus leads directly to the Lorentz combining rule of equation (41) for the Beattie-Bridgeman constant B_0.

The variational method which leads to equation (53) gives the following expression for the leading term of the attractive energy between two unlike atoms (Buckingham, 1937):

$$u_{12}(r) = -3E_0 a_0^{3/2} r^{-6} \frac{\alpha_1 \alpha_2}{\sqrt{(\alpha_1/n_1)} + \sqrt{(\alpha_2/n_2)}}. \qquad \ldots\ldots(74)$$

It is seen that if α_1/n_1 is not too different from α_2/n_2 (and it should not be, for the polarizability does increase with the number of electrons), a very good approximation can be made by taking u_{12} as the geometric mean of u_{11} and u_{22}. When this assumption is introduced into (59), and it is assumed that the sizes are not too different, the square-root combination rule for A_0, given by equation (40), is indicated.

Calculations of B_0 for mixtures, with the aid of assumptions corresponding to these rough theoretical considerations, have been made with the Lennard-Jones potential for mixtures of hydrogen and nitrogen (Lennard-Jones and Cook, 1927), helium and nitrogen (Hirschfelder and Roseveare, 1939), and methane and n-butane (Stockmayer, 1940).

In concluding this report on the theory, it is desired to leave the impression that the basic principles of the subject seem very firmly established, but that much difficult mathematics must be conquered before the goal of quantitative *a priori* prediction can be achieved.

REFERENCES

Amagat, E. H., 1893. *Ann. Chim.* (*Phys.*) **29**, 68, 505.
Baber, T. D. H. and Hassé, H. R., 1937. *Proc. Camb. Phil. Soc.* **33**, 253.
Band, W., 1939. *J. Chem. Phys.* **7**, 324, 927.
Barus, C., 1890. *Phil. Mag.* **30**, 338.
Beattie, J. A., 1924. *J. Amer. Chem. Soc.* **46**, 342.
Beattie, J. A., 1927. *J. Amer. Chem. Soc.* **49**, 1123.
Beattie, J. A., 1928 a. *Phys. Rev.* **31**, 680.
Beattie, J. A., 1928 b. *Phys. Rev.* **32**, 699.
Beattie, J. A., 1928 c. *Phys. Rev.* **32**, 691.
Beattie, J. A., 1929 a. *Phys. Rev.* **34**, 1615.
Beattie, J. A., 1929 b. *J. Amer. Chem. Soc.* **51**, 19.
Beattie, J. A., 1930 a. *Phys. Rev.* **36**, 132.
Beattie, J. A., 1930 b. *Proc. Nat. Acad. Sci., Wash.*, **16**, 14.
Beattie, J. A., 1930 c. *J. Math. Phys., M.I.T.*, **9**, 11.
Beattie, J. A., 1930 d. *Phys. Rev.* **35**, 643.
Beattie, J. A. and Bridgeman, O. C., 1927. *J. Amer. Chem. Soc.* **49**, 1665.
Beattie, J. A. and Bridgeman, O. C., 1928 a. *J. Amer. Chem. Soc.* **50**, 3133.
Beattie, J. A. and Bridgeman, O. C., 1928 b. *Proc. Amer. Acad. Arts Sci.* **63**, 229.
Beattie, J. A. and Bridgeman, O. C., 1928 c. *J. Amer. Chem. Soc.* **50**, 3151.
Beattie, J. A. and Bridgeman, O. C., 1930. *Z. Phys.* **62**, 95.

BEATTIE, J. A., HADLOCK, C. and POFFENBERGER, N., 1935. *J. Chem. Phys.* **3**, 93.
BEATTIE, J. A. and IKEHARA, S., 1930. *Proc. Amer. Acad. Arts Sci.* **64**, 127.
BEATTIE, J. A., INGERSOLL, H. G. and STOCKMAYER, W. H. To be published.
BEATTIE, J. A., KAY, W. C. and KAMINSKY, J., 1937. *J. Amer. Chem. Soc.* **59**, 1589.
BEATTIE, J. A. and LAWRENCE, C. K., 1930. *J. Amer. Chem. Soc.* **52**, 6.
BEATTIE, J. A., SIMARD, G. L. and SU, G.-J., 1939. *J. Amer. Chem. Soc.* **61**, 26.
BENEDICT, M., 1940. Private communication.
BENEDICT, M., WEBB, G. B. and RUBIN, L. C., 1940. *J. Chem. Phys.* **8**, 334.
BERTHELOT, D., 1898. *C.R. Acad. Sci., Paris*, **126**, 1703, 1857.
BERTHELOT, D., 1907. *Trav. Bur. Int. Poids Mes.* **13**.
BLEICK, W. E. and MAYER, J. E., 1934. *J. Chem. Phys.* **2**, 252.
BOER, J. DE and MICHELS, A., 1938. *Physica*, **5**, 945.
BOER, J. DE and MICHELS, A., 1939 a. *Physica*, **6**, 409.
BOER, J. DE and MICHELS, A., 1939 b. *Physica*, **6**, 97.
BORN, M., 1937. *Physica*, **4**, 1034.
BORN, M. and FUCHS, K., 1938. *Proc. Roy. Soc.* A, **166**, 391.
BRIDGEMAN, O. C., 1927. *J. Amer. Chem. Soc.*, **49**, 1130.
BRIDGEMAN, O. C., 1929. *Phys. Rev.* **34**, 527.
BROWN, G. G., SOUDERS, M., jr. and SMITH, R. L., 1932. *Industr. Engng Chem.* **24**, 513.
BUCKINGHAM, R. A., 1937. *Proc. Roy. Soc.* A, **160**, 113.
BUCKINGHAM, R. A., 1938. *Proc. Roy. Soc.* A, **168**, 264.
CALLENDAR, H., 1928. *Proc. Roy. Soc.* A, **120**, 460.
COPE, J. Q., LEWIS, W. K. and WEBER, H. C., 1931. *Industr. Engng Chem.* **23**, 887.
CORNER, J., 1939. *Trans. Faraday Soc.* **35**, 784.
DEMING, W. E. and SHUPE, L. E., 1929. *J. Franklin Inst.* **208**, 389.
DEMING, W. E. and SHUPE, L. E., 1931 a. *Phys. Rev.* **37**, 638.
DEMING, W. E. and SHUPE, L. E., 1931 b. *Phys. Rev.* **38**, 2245.
DEMING, W. E. and SHUPE, L. E., 1932. *Phys. Rev.* **40**, 848.
DIETERICI, C., 1899. *Ann. Phys., Lpz.*, **69**, 685.
DIETERICI, C., 1901. *Ann. Phys., Lpz.*, **5**, 51.
DONDER, TH. DE, 1925. *C.R. Acad. Sci., Paris*, **180**, 1922.
DONDER, TH. DE and RYSSELBERGHE, P. VAN, 1936. *Thermodynamic Theory of Affinity* (Stanford University Press, Calif.).
EISENSCHITZ, R., and LONDON, F., 1930. *Z. Phys.* **60**, 491.
EUCKEN, A. and BRESLER, F., 1928. *Z. phys. Chem.* **134** A, 230.
FOWLER, R. H., 1936 a. *Statistical Mechanics*, 2nd edition (Cambridge), Chap. IX, p. 286.
FOWLER, R. H., 1936 b. *Statistical Mechanics*, 2nd edition (Cambridge), Chap. VIII, p. 242.
FOWLER, R. H., 1936 c. *Statistical Mechanics*, 2nd edition (Cambridge), Chap. X.
FOWLER, R. H. and GUGGENHEIM, E. A., 1939. *Statistical Thermodynamics*, Chap. VII.
FRENKEL, J., 1939. *J. Chem. Phys.* **7**, 200.
GALITZINE, B., 1890. *Ann. Phys., Lpz.*, **41**, 770.
GIBSON, G. E. and SOSNICK, B., 1927. *J. Amer. Chem. Soc.* **49**, 2172.
GILLESPIE, L. J., 1925. *J. Amer. Chem. Soc.* **47**, 305, 3106.
GILLESPIE, L. J., 1926. *J. Amer. Chem. Soc.* **48**, 28.
GILLESPIE, L. J., 1929 a. *J. Phys. Chem.* **33**, 354.
GILLESPIE, L. J., 1929 b. *Phys. Rev.* **34**, 352, 1605.
GILLESPIE, L. J., 1936. *Chem. Rev.* **18**, 359.
GILLESPIE, L. J. and BEATTIE, J. A., 1930 a. *Phys. Rev.* **36**, 743, 1008.
GILLESPIE, L. J. and BEATTIE, J. A., 1930 b. *J. Amer. Chem. Soc.* **52**, 4239.
GOMBÁS, P., 1935. *Z. Phys.* **93**, 378.
GROPPER, L., 1937. *Phys. Rev.* **51**, 50, 1108.
HAPPEL, H., 1906. *Ann. Phys., Lpz.*, **21**, 342.
HARRISON, S. F. and MAYER, J. E., 1938. *J. Chem. Phys.* **6**, 101.

HASSÉ, H. R., 1930. *Proc. Camb. Phil. Soc.* **26**, 542.
HASSÉ, H. R., 1931. *Proc. Camb. Phil. Soc.* **27**, 66.
HEITLER, W. and LONDON, F., 1927. *Z. Phys.* **44**, 455.
HELLMANN, H., 1935. *Acta Physicochimica, U.S.S.R.*, **2**, 273.
HIRSCHFELDER, J. O. and ROSEVEARE, W. E., 1939. *J. Phys. Chem.* **43**, 15.
JENSEN, H., 1936. *Z. Phys.* **101**, 164.
KAHN, B., 1938. *Dissertation* (Utrecht).
KAHN, B. and UHLENBECK, G. E., 1938. *Physica*, **5**, 399.
KEESOM, W. H., 1912. *Commun. Phys. Lab. Leiden*, Suppl. no. 24 B, 32.
KENNARD, E. H., 1938. *Kinetic Theory of Gases* (McGraw-Hill), Chap. v, p. 225.
KEYES, F. G., 1914. *J. Amer. Soc. Refrig. Engrs*, **1**, 9.
KEYES, F. G., 1917. *Proc. Nat. Acad. Sci., Wash.*, **3**, 323.
KEYES, F. G., 1920. *J. Amer. Chem. Soc.* **42**, 54.
KEYES, F. G., 1921 a. *J. Amer. Soc. Refrig. Engrs*, **7**, 371.
KEYES, F. G., 1921 b. *J. Amer. Chem. Soc.* **43**, 1452.
KEYES, F. G., 1922. *J. Math. Phys., M.I.T.*, **1**, 89.
KEYES, F. G., 1924. *J. Amer. Chem. Soc.* **46**, 1584.
KEYES, F. G., 1927. *J. Amer. Chem. Soc.* **49**, 1393.
KEYES, F. G., 1929. *Chem. Rev.* **6**, 175.
KEYES, F. G., 1938. *J. Amer. Chem. Soc.* **60**, 1761.
KEYES, F. G. and BURKS, H. G., 1927. *J. Amer. Chem. Soc.* **49**, 1403.
KEYES, F. G. and BURKS, H. G., 1928. *J. Amer. Chem. Soc.* **50**, 1100.
KEYES, F. G. and FELSING, W. A., 1919. *J. Amer. Chem. Soc.* **41**, 589.
KEYES, F. G. and FELSING, W. A., 1920. *J. Amer. Chem. Soc.* **42**, 106.
KEYES, F. G. and KENNEY, A. W., 1917. *J. Amer. Soc. Refrig. Engrs*, **3**, 1.
KEYES, F. G. and SEARS, F. W., 1925. *Proc. Nat. Acad. Sci., Wash.*, **11**, 38.
KEYES, F. G., SMITH, L. B. and GERRY, H. T., 1934. *Proc. Amer. Acad. Arts Sci.* **69**, 137, 285.
KEYES, F. G., SMITH, L. B. and GERRY, H. T., 1936. *Proc. Amer. Acad. Arts Sci.* **70**, 319.
KEYES, F. G., SMITH, L. B., JOUBERT, D. B. and TAYLOR, R. S., 1922. *J. Math. Phys., M.I.T.*, **1**, 191, 211.
KEYES, F. G. and TAYLOR, R. S., 1927. *J. Amer. Chem. Soc.* **49**, 896.
KEYES, F. G., TOWNSEND, B. and YOUNG, L. H., 1922. *J. Math. Phys., M.I.T.*, **1**, 243.
KIRKWOOD, J. G., 1932. *Phys. Z.* **33**, 57.
KIRKWOOD, J. G., 1933. *J. Chem. Phys.* **1**, 597.
KIRKWOOD, J. G. and KEYES, F. G., 1931. *Phys. Rev.* **37**, 832.
KLEEMAN, R. D., 1928. *J. Franklin Inst.* **206**, 511.
KUENEN, J. P., 1907. *Die Zustandsgleichung der Gase und Flüssigkeiten und die Kontinuitätstheorie* (Vieweg und Sohn, Braunschweig).
VAN LAAR, J. J. and LORENZ, R., 1925. *Z. anorg. Chem.* **145**, 239.
VAN LAAR, J. J., 1928 a. *Z. anorg. Chem.* **171**, 42.
VAN LAAR, J. J., 1928 b. *Z. phys. Chem.* **137** A, 421.
VAN LAAR, J. J., 1929. *Z. phys. Chem.* **145** A, 207.
LASHAKOV, L., 1939. *Acta Physicochimica, U.S.S.R.*, **11**, 107.
LENNARD-JONES, J. E., 1924. *Proc. Roy. Soc.* A, **106**, 463.
LENNARD-JONES, J. E., 1931. *Proc. Phys. Soc.* **43**, 461.
LENNARD-JONES, J. E., 1937. *Physica*, **4**, 941.
LENNARD-JONES, J. E. and COOK, W. R., 1927. *Proc. Roy. Soc.* A, **115**, 334.
LENNARD-JONES, J. E. and DEVONSHIRE, A. F., 1937. *Proc. Roy. Soc.* A, **163**, 53
LENNARD-JONES, J. E. and DEVONSHIRE, A. F., 1938. *Proc. Roy. Soc.* A, **165**, 1.
LENNARD-JONES, J. E. and DEVONSHIRE, A. F., 1939. *Proc. Roy. Soc.* A, **169**, 317.
LERBERGHE, G. VAN, 1925. *C.R. Acad. Sci., Paris*, **181**, 851.
LERBERGHE, G. VAN, 1928. *Bull. Acad. Belg. Sci.* **14**, 349.
LEWIS, W. K., 1936. *Industr. Engng Chem.* **28**, 257.

LEWIS, W. K. and LUKE, C. D., 1933. *Industr. Engng Chem.* **25**, 725.
LEWIS, G. N. and RANDALL, M., 1923. *Thermodynamics* (McGraw-Hill Book Co., New York).
LONDON, F., 1930 a. *Z. phys. Chem.* B, **11**, 222.
LONDON, F., 1930 b. *Z. Phys.* **63**, 245.
LONDON, F., 1937. *Trans. Faraday Soc.* **33**, 8.
LORENTZ, H. A., 1881. *Ann. Phys., Lpz.*, **12**, 127, 660.
LURIE, E. and GILLESPIE, L. J., 1927. *J. Amer. Chem. Soc.* **49**, 1146.
MAASS, O., 1938. *Chem. Rev.* **23**, 17.
MACDOUGALL, F. H., 1916. *J. Amer. Chem. Soc.* **38**, 528.
MACDOUGALL, F. H., 1917. *J. Amer. Chem. Soc.* **39**, 1229.
MAMOTENKO, M. F., 1939. *Acta Physicochimica, U.S.S.R.*, **11**, 225.
MAMOTENKO, M. F. and SCHUCHOWITZKY, A., 1938. *Acta Physicochimica, U.S.S.R.*, **9**, 803.
MARGENAU, H., 1939 a. *Phys. Rev.* **56**, 1000.
MARGENAU, H., 1939 b. *Rev. Mod. Phys.* **11**, 1.
MASSEY, H. S. W. and BUCKINGHAM, R. A., 1938 a. *Min. Proc. R. Irish Acad.* A, **45**, 31.
MASSEY, H. S. W. and BUCKINGHAM, R. A., 1938 b. *Proc. Roy. Soc.* A, **168**, 378.
MAYER, J. E., 1937. *J. Chem. Phys.* **5**, 67.
MAYER, J. E., 1939. *J. Phys. Chem.* **43**, 71.
MAYER, J. E. and ACKERMANN, P. G., 1937. *J. Chem. Phys.* **5**, 74.
MAYER, J. E. and GOEPPERT-MAYER, M., 1940. *Statistical Mechanics* (New York: John Wiley), Chaps. XIII & XIV.
MAYER, J. E. and HARRISON, S. F., 1938. *J. Chem. Phys.* **6**, 87.
MAYER, J. E. and STREETER, S. F., 1939. *J. Chem. Phys.* **7**, 1025.
MERZ, A. R. and WHITTAKER, C. W., 1928. *J. Amer. Chem. Soc.* **50**, 1522.
NEWTON, R. H., 1935. *Industr. Engng Chem.* **27**, 302.
NEWTON, R. H. and DODGE, B. F., 1935. *Industr. Engng Chem.* **27**, 577.
ONNES, H. KAMERLINGH, 1901. *Commun. Phys. Lab. Univ. Leiden*, no. 71.
ONNES, H. KAMERLINGH and KEESOM, W. H., 1912. *Commun. Phys. Lab. Univ. Leiden*, Supplement no. 23.
OTTO, J., 1929. *Hand. Experimentalphysik*, pp. 137–180 (Akademische Verlagsgesellschaft M.B.H., Leipzig).
PAULING, L. and BEACH, J. Y., 1935. *Phys. Rev.* **47**, 686.
PHILLIPS, H. B., 1921. *Proc. Nat. Acad. Sci., Wash.*, **7**, 172.
RAMSAY, W. and YOUNG, S., 1887 a. *Phil. Mag.* **23**, 435.
RAMSAY, W. and YOUNG, S., 1887 b. *Z. phys. Chem.* **1**, 237, 433.
RANDALL, M. and SOSNICK, B., 1928. *J. Amer. Chem. Soc.* **50**, 967.
SAGE, B. H. and LACY, W. N., 1939. *Industr. Engng Chem.* **31**, 1497.
SAGE, B. H., WEBSTER, D. C. and LACY, W. N., 1937. *Industr. Engng Chem.* **29**, 658, 1188.
SCATCHARD, G., 1930. *Proc. Nat. Acad. Sci., Wash.*, **16**, 811.
SCHÄFER, K., 1937 a. *Z. phys. Chem.* B, **36**, 85.
SCHÄFER, K., 1937 b. *Z. phys. Chem.* B, **38**, 187.
SLATER, J. C., 1928. *Phys. Rev.* **32**, 349.
SLATER, J. C. and KIRKWOOD, J. G., 1931. *Phys. Rev.* **37**, 682.
SMITH, L. B., BEATTIE, J. A. and KAY, W. C., 1937. *J. Amer. Chem. Soc.* **59**, 1587.
SMITH, L. B. and TAYLOR, R. S., 1923 a. *J. Amer. Chem. Soc.* **45**, 2107.
SMITH, L. B. and TAYLOR, R. S., 1923 b. *J. Amer. Chem. Soc.* **45**, 2124.
STOCKMAYER, W. H. To be published.
SUGIURA, Y., 1927. *Z. Phys.* **45**, 484.
TAPP, J., STEACIE, E. W. and MAASS, O., 1933. *Canad. J. Res.* **9**, 217.
THOMSON, J., 1871. *Proc. Roy. Soc.* A, **20**, 1.
TRAUTZ, M. and EMERT, O., 1926. *Z. anorg. Chem.* **150**, 277.

TRAUTZ, M. and NARATH, A., 1926. *Ann. Phys., Lpz.*, **79**, 637.
TSENG, T. P., FENG, S. K., CHENG, C. and BAND, W., 1940. *J. Chem. Phys.* **8**, 20.
UHLENBECK, G. E. and BETH, E., 1936. *Physica*, **3**, 729.
UHLENBECK, G. E. and BETH, E., 1937. *Physica*, **4**, 915.
URSELL, H. D., 1927. *Proc. Camb. Phil. Soc.* **23**, 685.
VERSCHOYLE, T. T. H., 1926. *Proc. Roy. Soc.* A, **111**, 552.
WAALS, J. D. VAN DER, 1899 and 1900. *Die Continuität des Gasförmigen und Flüssigen Zustandes* (Barth, Leipzig).
YOUNG, S., 1895. *Proc. Phys. Soc.* **13**, 602.

ELECTRICAL DISCHARGES IN GASES AND THEIR APPLICATIONS : PART I*

By V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P., A.M.I.E.E. and H. G. JENKINS, M.Sc., F.Inst.P.

(Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England)

§1. INTRODUCTION

Although little that is fundamentally new has arisen in the study of electric discharges through gases in recent years, substantial advances have been made in the understanding of many phenomena. Because the subject provides an unrivalled field for experiment and a wealth of complicated and interesting processes to be sorted out, a vast literature continues to be published. This has been stimulated in the last few years by the increasing commercial importance of discharge devices.

In a short review of so extensive a branch of physics, some selection is inevitable and no attempt has been made to cover the whole subject. Those phenomena have been included which have recently become less obscure either from a descriptive point of view, such as the long spark in air, or from a theoretical point of view, such as the mechanism of plasma. The other aspect of the subject which has been considered of outstanding importance is the recent advance in the attempts to explain discharge phenomena in terms of atomic processes. Good examples of this are the calculations of electron velocity distribution in plasma and the calculation of electron mobility. It has been thought worth while to include a large number of recent references, even when space does not permit these to be more than mentioned.

The processes taking place in a gas discharge may be roughly classified into three groups:

(*a*) Individual processes.
(*b*) Average processes.
(*c*) Macroscopic phenomena.

Under (*a*) can be considered the individual acts of ionization and excitation by ions and atoms, the action of photons and ions at surfaces, and similar processes. This group is largely a branch of atomic physics. The average processes usually involve several individual processes; for example, diffusion, mobility and

* As a result of the present conditions, it has been found impossible to prepare Part II, dealing with the practical applications, in time for this volume of the *Progress Reports*. It is hoped that Part II will appear in the next volume.

Townsend's ionization coefficients are all dependent on the parameters of several individual processes. Similarly, the macroscopic phenomena, such as the positive column and the negative glow, involve all the average processes to a greater or lesser extent.

The logical and ideal approach to the study of discharges would be to obtain the necessary constants for the individual processes, and to deduce from these a full knowledge of the average processes, which could then be applied to interpret the macroscopic phenomena. In the earlier days of gas-discharge physics, since very little was known about the individual processes, the tendency was in the opposite direction, that is to obtain information about the individual processes from the average and macroscopic. This method of obtaining quantitative information about atomic processes, however, has never been very successful.

With the gradually increasing data on atomic processes in recent years, this tendency has been reversed. The problem, however, is by no means simple. In the first place there are so many individual processes involved in any discharge that, even if these are understood and experimental methods are available for their quantitative study, the amount of experimental work is very formidable. In the second place, even with full information of all the necessary parameters, the mathematical difficulties involved in applying it to quantitative calculation of average processes are often prohibitive. Nevertheless, the most notable advance in the last ten years or so has been in reducing this gap which exists between the study of the individual processes and the quantitative interpretation of discharge phenomena in terms of them. In some cases, as will be mentioned below, progress in this direction has been very successful.

The present position may be roughly summarized as follows:—Nearly all the individual processes are understood qualitatively; in most cases, semi-quantitative data are available and in a few cases reliable values of parameters are known. With regard to the average processes, again, in nearly all cases their qualitative interpretation in terms of the individual processes is well established; where the parameters of the individual processes are available, only the tedious mathematics involved prevent their complete calculation. The macroscopic phenomena, such as the study of the positive column and negative glow, are, in general, believed to be well understood qualitatively; it is only to be expected, however, that comprising as they do much more complicated sets of phenomena than any single average process, the theoretical development is not so far advanced.

This is a much more encouraging position than that existing even a few years ago. There is, however, much yet to be done before the study and development of discharge devices can be carried out on anything other than a purely empirical basis. A somewhat unsatisfactory state of affairs in this respect is to be expected since, of all physical phenomena, electrical discharges in gases are probably the most complicated.

In this review no more can be done than to discuss briefly the present position in connection with the various processes dealt with and to indicate the lines along

which recent work has been proceeding. There are several recent books and reviews which should be consulted for further information: Arnot's book on collision processes in gases (1933) deals with most of the important individual processes up to 1932, and Brode (1933) has also reviewed the data on electron collisions; Loeb (1936) has discussed the mechanism of the static spark discharge; Massey (1938) has dealt with the recent work on negative ions, Tyndall (1938) with the recent study of the mobility of positive ions in gases. Loeb's book (1939 a) gives a critical survey in detail of the present position of gas-discharge processes, and frequent reference has been made to this work in writing the present review; Uyterhoeven's book (1938) treats the subject with particular reference to luminous-discharge lamps.

§ 2. INDIVIDUAL AND COLLISION PROCESSES

The study of collision processes is really a branch of atomic physics. It is of fundamental importance for the theory of gas discharges, however, since ultimately all the macroscopic phenomena of the discharge must be explainable in terms of the constants of the atomic processes.

There is a certain probability of any particular process (e.g. elastic collision, ionization) taking place at an impact between two particles. This probability may be conveniently expressed in terms of an effective collision cross-section for that process. Thus, if Q is the effective cross-section for the type of collision being considered, the number of collisions made by a particle in travelling 1 cm. through the gas which result in the process taking place is nQ, where n is the number of atoms per cm^3. The product nQ is called the " efficiency " for the process considered and is usually reduced to its value at 1 mm. pressure and 0° c. If the total collision cross-section is ΣQ, the probability of the process taking place at a collision is $Q/\Sigma Q$. The data required in a study of gas discharges are embodied in the {effective cross-section *vs.* velocity of particle} curves for the various types of collision processes, that is, for example, in the ionization and excitation functions. For a given velocity of the particle (electron, ion, atom or photon) and a given target molecule,* it is necessary to know not only the total cross-section for collision but also the cross-section for ionization and for excitation to the various levels and, in the case of collisions of the second kind, the cross-sections for the various types of energy transfer that are possible.

(a) *Electron collisions*

The most important individual process is that of ionization by electron impact, since upon this process depends the supply of electrons to allow a discharge to be self-maintained. Excitation and elastic scattering, while not playing quite such a vital part, to a large extent control the velocity distribution and mobility of the electrons and determine the energy balance in the discharge. Reliable figures for most ionization potentials and excitation potentials of importance in gas discharges are available from spectroscopic measurements, and experiments

* The term " molecules " will, in this review, include the atoms of monatomic gases.

to obtain these values directly continue to be made (see table 6*). Data on the ionization and excitation functions, however, are not by any means complete; the position is better in the case of ionization than in the case of excitation.

It is usual to measure the number of electrons (N) produced by one electron passing through 1 cm. of the gas at 1 mm. Hg pressure at 0° c., rather than the

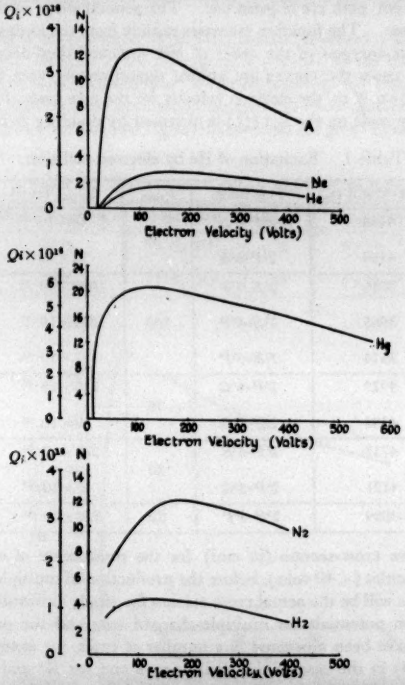

Figure 1. Efficiency of ionization (N) and effective cross-section for ionization (Q_i).

actual cross-section for ionization. To obtain the effective cross-section for single ionization from this figure it is necessary to know the numbers of the various multiple-charged ions which are formed. When the number of multiple-charged ions is subtracted from the total number of ions to give the number of singly-charged ions, the effective cross-section for single ionization is obtained by

* A number of references are given in table 6 which are not necessarily mentioned in the text.

dividing this number by the number of atoms per cm.^{3} ($3{\cdot}56 \times 10^{16}$). In actual fact, in studying processes in discharges, the electron velocities are not usually high enough to make the formation of multiple-charged ions very important; and, in any case, since what is generally required in practice is to know the number of electrons produced, the curves giving the total number of electrons produced per electron per cm. path are of great use. The general shape of these curves is similar for all gases. The function increases rapidly from the ionization potential to a maximum at energies of the order of 100 volts and then decreases slowly; in a number of cases the curves are known quantitatively with fair accuracy. In figure 1 is shown N *vs.* the electron velocity for the rare gases, for Hg and for N_2 and H_2. The scale on the left (Q_i) is obtained by dividing N by $3{\cdot}56 \times 10^{16}$,

Table 1. Excitation of He by electron collision

Wave-length (A.)	Transition	V_{max}	$Q_{e\ max}$ (cm^2)
4438	2^1P–5^1S	36	28×10^{-21}
4169	2^1P–6^1S		12×10^{-21}
5016	2^1S–3^1P	100	1000×10^{-21}
3965	2^1S–4^1P		340×10^{-21}
3614	2^1S–5^1P		57×10^{-21}
4922	2^1P–4^1D	74	110×10^{-21}
4388	2^1P–5^1D		60×10^{-21}
4713	2^3P–4^3S	30	210×10^{-21}
4121	2^3P–5^3S		45×10^{-21}
3889	2^3S–3^3P	32	720×10^{-21}

and is an effective cross-section (in cm^2) for the production of one electron. At the lower velocities (<40 volts), before the production of multiple ions occurs to any extent, this will be the actual cross-section for single ionization.

The ionization potentials for multiple-charged ions and the probability of their formation have been measured in a number of cases, for example by Tate and Smith (1934) in the case of the alkali metals and for Kr and Xe, and by Bleakney (1930a, 1930c) for Hg, Ne and A. The efficiency of their formation is usually only a few per cent of that of singly charged ions up to several hundred volts energy, and they are not produced at all at the lower energies.

The determination of the excitation functions is much more difficult, and up to the present there are very few reliable quantitative data. The shape of the excitation-function curves is, however, similar to that of the ionization-function curves. They rise rapidly from the excitation potential to a maximum and then decrease more slowly. The position of the optimum value for excitation, relative

to the excitation potential, and the sharpness of the maximum depend upon the transition concerned. For example, in the case of excitation to triplet levels, the peak is much nearer to the excitation potential and the maximum is much sharper than for excitation to singlet levels. The effective cross-section for excitation (Q_e) varies widely according to the gas and type of level excited, being in some cases as high as 0·5 and in others lower than 0·001 of the total collision cross-section.

To show the order of magnitude of the quantities involved, a few measurements are given here. In table 1 are figures for a number of He lines (Lees, 1932). In the third and fourth columns are given respectively the voltage at which the ionization efficiency is a maximum and the effective cross-section at this voltage.

Similar figures for the two Hg resonance lines are given in table 2 (Arnot and Baines, 1935).

Table 2. Excitation of Hg by electron collision

Wave-length (A.)	Transition	V_{max}	$Q_{e\,max}$ (cm^2)
2536	1^1S_0–2^3P_1	8	4×10^{-16}
1850	1^1S_0–2^1P_1	16	4×10^{-16}

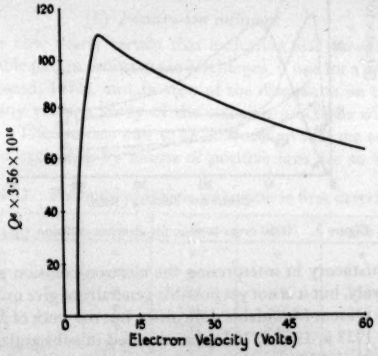

Figure 2. Excitation of the Na D-lines by electron collision.

In figure 2 is shown the excitation function for the Na D-lines (Haft, 1933; Christoph, 1935); the efficiency of excitation (nQ_e) is plotted against the electron velocity.

The total cross-section for collision on which the electron mean-free-path depends will include the effects of exciting and ionizing collisions as well as of elastic collisions. The collision cross-section is, of course, a function of the

electron velocity even in the case of elastic collision. The results of the earlier measurements on the total cross-section have been summarized by Kollath (1930 a). The approximate values for a few gases are shown in figure 3.

A knowledge of the mean free-paths is clearly necessary to any study of the motion of electrons through gases. The total cross-sections alone, however, are not sufficient to determine the motion, since the energy lost at each collision by an electron is an essential factor. For low-velocity electrons the collisions are mainly elastic, but as the electron velocity is increased the inelastic collisions become important and data on the effective cross-sections for the various possible energy losses at inelastic collisions become necessary. To obtain the probability curve for the total energy loss, it is necessary to add together the effective cross-sections at each electron velocity for all possible energy transfers.

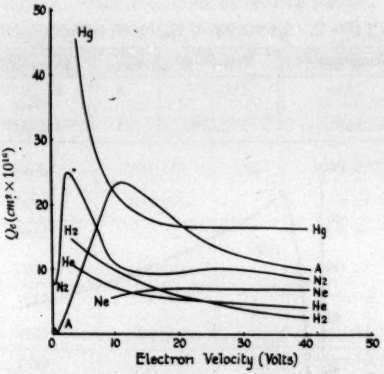

Figure 3. Total cross-section for electron collision (Q_e).

The status of theory in interpreting the electron-collision processes is satisfactory qualitatively, but it is not yet possible generally to give quantitative results. The calculations present formidable difficulties, but the work of Massey and Mohr (1931 a, 1931 b, 1933 a, 1933 b, 1934) has resulted in substantial progress in this direction. Penney (1932) has been able to explain theoretically the marked difference in the sharpness of the peaks in the excitation functions for singlet and triplet terms in the case of Hg vapour.

Much of the more recent experimental work in connection with electron collisions has been done on Hg vapour. Arnot and Baines (1935) have measured the excitation functions for the $2\,^1P_1$, $2\,^3P_1$ and $3\,^1D_2$ states of Hg (table 2) and the elastic, inelastic and total collision cross-sections. For electron velocities greater than about 10 volts, the elastic cross-section decreases rapidly, nearly the whole

of the effective cross-section being due to inelastic collisions. Snavely (1937) has studied experimentally the various processes taking place to produce cumulative ionization in Hg vapour. The results suggest that, for electron energies greater than 7 volts, ionization of metastable atoms accounts for the major part of the ionization. For lower electron energies it is suggested that the simultaneous formation of positive and negative ions as a result of collisions between excited atoms takes place. In a very interesting recent paper Nottingham (1939), following some preliminary work of Bell (1939), has measured the ionization function of Hg with sufficient refinement of technique to allow the detailed structure to appear. Data on the excitation functions of the $2\,^3P_1$, $3\,^3P_1$, and $4\,^3P_1$ levels were also obtained. Führmann (1939) has recently measured the excitation functions for several lines in the vapours of Pb, Tl and Ag.

With regard to ionization and excitation of molecular gases, a series of papers by Savard, Hemptinne and Capron (1933 *et seq.*) and by Löhner (1935 b) have appeared in the last few years in which a large number of non-polar molecular gases are studied. It is perhaps fortunate that molecular gases are not of the same practical importance in gas discharges as monatomic gases, since the collision processes are so complicated that it will be a long time before the required quantitative data are available.

(b) *Positive-ion collisions*

Although it is now nearly certain that ionization and excitation by positive ions play a negligible part in ordinary gas discharges, it was for a long time thought otherwise (Townsend, 1915), and, in view of the discussion on this topic, which has lasted for many years, a study of the collision processes with positive ions is of importance. There seems now to be no question that the energies required for ionization and excitation by means of positive ions are so high as to make

Table 3. Potential at which ionization is first detected

	Ne	A	Kr	Xe
Li^+	•	•	•	•
Na^+	130	•	•	•
K^+	•	82	69	114
Rb^+	•	135	97·5	150
Cs^+	•	338	200	77
Electrons	21·5	15·5	·13	11·5

* No ionization detected up to the highest energy used (apparently not stated in Varney's paper).

these processes very improbable. The greatest interest from a practical viewpoint lies in the collisions of positive ions with their own atoms. Experimentally, these data are difficult to obtain, and there are few results available. It seems fairly clear, however, that appreciable ionization does not occur with positive-ion energies less than several hundred volts (see, for example, Bracefield, 1932).

The position is rather different in the case of ionization of rare-gas atoms by positive ions of the alkali-metal vapours. Varney (1935 a), using an improved technique, has measured the potentials required to ionize Ne, A, Kr, and Xe by positive ions of Li, Na, Rb and Cs. His results are given in table 3 together with the normal ionization potentials of the atoms for comparison.

These results are in substantial agreement with other measurements (see table 6) except that in all cases they are somewhat lower (to the extent of 10 to 40 volts). The reason for the relatively low potentials in the case of ionization of rare-gas atoms by positive ions of the alkali metals does not seem to be well understood. It is probably a resonance effect associated with the electronic structures of

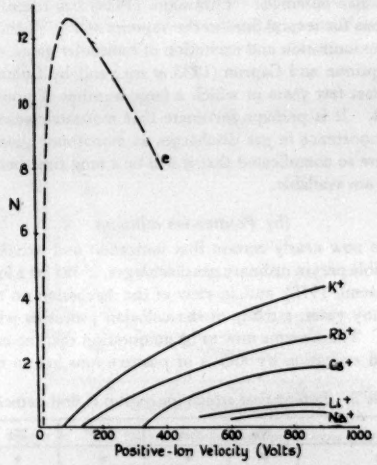

Figure 4. Efficiency of ionization (N) for A bombarded by alkali positive ions.

the atom and ion, and seems to be most efficient when these structures are identical. There are, however, some puzzling results. It has been found, for example, that Na^+ will ionize Hg at 88 volts, but that K^+ will not ionize Hg at energies up to 300 volts (Varney *et al.*, 1936 b, 1937).

Even in these cases, however, the most favourable for ionization by positive ions, the energies required are at least five times those necessary for ionization by electrons. There seems to be no case on record in which ionization of atoms by positive ions has been observed at energies less than about 50 volts, and most combinations of gases and ions require energies of the order of several hundred volts. Even when ionization occurs, the efficiency is, as would be expected, low

near the onset potential. There are few data on the shape of the ionization-function curves except in the case of rare-gas atoms bombarded by alkali ions. At voltages well above the onset voltage, the ionization efficiency is quite high. In figure 4 are shown the efficiency-of-ionization curves for A bombarded by alkali positive ions. On the same scale, for comparison, is shown the electron-ionization function from figure 1. References from which further data can be obtained are given in table 6.

Maurer (1939) has recently reviewed the available data on excitation. It seems well established that positive ions of energy less than that required for ionization will excite their own or other atoms. The energies required, however, are still far greater than those normally possessed by positive ions in discharges. A number of measurements have been made on excitation functions, references to which are given in table 6.

With regard to the total cross-section for collision, which determines the mean free-path, except for high energies the cross-sections for elastic scattering and for neutralization predominate. This is distinct from the case of electron impact, where, even at low electron energies, inelastic collisions have to be considered. Beeck (1934), Schmidt (1934 a, 1934 b, 1936), and Bracewell (1938) have recently made measurements on elastic scattering, and Massey and Smith (1933 c) have treated theoretically the problem of the passage of positive ions through gases. The collision cross-sections are functions of the positive-ion velocity and frequently have a minimum at an energy of the order of 100 volts, increasing rapidly as the velocity of the positive ion is decreased. The cross-section for neutralization is often very large, particularly when the ions are passing through a gas of similar atoms.

In normal discharges, ionization and excitation by positive ions are usually negligible, since the positive-ion energies are rarely in excess of a few volts. The main importance of the collision processes is therefore in connection with the mean free-path and mobility and the way in which the space-charge distribution is influenced thereby. As in the case of electron collisions, it is not sufficient to know merely the mean free-path of the ions to deduce their motion through the gas. In addition, the probability of the various types of collision such as scattering and electron transfer must be known, and the energy loss at these different types of collisions as a function of the ion velocity must also be known. Data are not yet available to allow quantitative estimates of these effects to be made. The result of this lack of data is seen, for example, in the difficulty in interpreting some of the results of the positive-ion mobility measurements in gases.

(c) *Collisions with atoms*

The neutral atoms in a gas discharge for the most part have an energy of only a small fraction of a volt. A few atoms may, however, have somewhat higher energies as the result of their formation by the capture of electrons by positive ions. Although it appears that ionization of atoms by similar atoms is a more

efficient process than ionization by positive ions, these energies are still far too low to be of any importance in most discharges. An exception to this probably occurs in high-temperature arcs at high pressure, when, under conditions of thermal ionization, the atomic-collision processes may be of considerable importance.

Bracefield (1933) and Varney (1936 a) found ionization in the rare gases bombarded by their own atoms with energies of 100 volts or less. No ionization could be detected in atoms bombarded by other atoms up to several hundred volts energy. Wayland (1937), however, obtained evidence of ionization in the case of bombardment by dissimilar atoms. This question is probably associated with the resonance effect already mentioned in connection with the ionization of atoms by positive ions. Varney's results are given in table 4, with Wayland's results in parenthesis.

Table 4. Potential at which ionization is first detected

	Ne	A	Kr	Xe
Ne	(74)			
A	130	48 (48)	50	55
Kr			(40)	
Xe				(35)

Whether or not ionization of one gas by neutral atoms of another gas does sometimes take place at a few tens of volts of energy, or only at much higher values, is thus not clear. In any case, there is no doubt that the energies required are far greater than those normally occurring in gas discharges.

There has been a fair amount of work on excitation by neutral-atom impact, but nearly all of the work is at energies far greater than those of interest in this connection; there has been little work on the detection of excitation at low energies. Maurer's review (1939) gives a useful summary of the position to date.

(d) *Photo-excitation and Photo-ionization*

Excitation and ionization by radiation are of great importance in discharge mechanism. In some cases they may even be the major processes operative (for example, in the case of the propagation of long sparks at atmospheric pressure). In other cases, the absorption and re-emission of light may profoundly modify discharge-tube characteristics, particularly in gas mixtures or when impurities are present.

Cravath (1935 a) appears to have proved that the short-wave radiation produced in discharges is capable of acting as an effective secondary process in sparking in air at atmospheric pressure. Varney and Loeb (1935 c), in this same connection, have studied the photo-ionization produced in N_2, A and Xe under various conditions, and conclude that there is only a very small probability of ionization occurring by the absorption of two quanta, each having only a fraction of the necessary energy. Raether (1938) has carried out an interesting experiment

in which he measures the ionization in air, H_2 and O_2 when irradiated with the light from a spark. It is possible that developments of experiments on these lines would produce some much-needed data. Most of the work that has been done recently on photo-ionization, however, has been in the x-ray region, and is of little direct interest for the study of gas discharges. Experiments in the useful range of wave-lengths on the excitation and ionization functions involve considerable and well known difficulties. The general trend of the curves in some cases is known (see Arnot, 1933), but quantitative data are very few.

A good deal of work continues to be carried out on various other aspects of photo-excitation in gases (Hanle, 1934, 1938), particularly in connection with the mean lives of excited states, and some useful data have resulted. There is nothing fundamentally new, however, nor are most of the data in a form in which they can be used directly for the interpretation of discharge phenomena.

(e) *Collisions of the second kind*

Under this heading may be included all those processes in which energy is transferred during a collision from an excited atom to another particle, electron or molecule. A special case is the collision between an ion and a normal molecule in which the ion returns to a normal molecule and the normal molecule is ionized. This involves the transfer of an electron as well as of energy. All the recent work, both experimental and theoretical, agrees in assigning an important rôle to these processes, particularly when the excited atom is in a metastable state. Penning (1931 a, 1932), for example, has shown that the sparking potential of a gas may be materially affected by the presence of metastable atoms. There is now no doubt that in nearly all gaseous-discharge phenomena, this type of energy exchange takes place, sometimes with very high probability (see Kenty, 1933; Ferchmin *et al.*, 1936; Spiwak *et al.*, 1936).

The effectiveness of metastable atoms lies partly in their long lives, as a result of which they are able to store energy with a high probability of passing it on to another particle, and partly in the fact that, since they are uncharged, they are able by diffusion to carry their energy against the fields into parts of the discharge from which they might otherwise be excluded.

In view of their importance, considerable attention has been paid to collisions of these types during the last few years, although again the position is that data on the cross-sections as a function of velocity of the colliding particles are very few. The probability of an energy transfer depends largely on the difference between the energy available and that required for ionization or excitation, being a maximum when there is exact resonance. It seems that the effective cross-section at this maximum may be very much higher than the gas-kinetic cross-section. There are, however, other factors involved, such as the type of level to be excited, which help to determine the effective collision cross-section for energy transfer (cf. Duffendack *et al.*, 1933).

As examples of the experimental technique and the type of measurements which are being made, the following papers may be referred to:—Manley and

Duffendack (1935), by examining the line intensity in the spectrum of Mg II in a mixture of Mg vapour and Ne, have deduced a value of about 10^{-11} cm^2 for the cross-section for collisions in the case of small energy discrepancy between Ne^+ and Mg atoms. Duffendack and Gran (1937) have obtained an effective cross-section of the same order for collisions of Ne^+ with Pb atoms. The quenching of Hg fluorescence by H_2, CO and N_2 has been studied by Duffendack and Owens (1934), and that of Cd fluorescence by Lipson and Mitchell (1935). Kruithof and Druyvesteyn (1937 b), using the known value of the ionization coefficient for Ne-A mixtures, have estimated the probability of ionization of A atoms in a discharge by metastable Ne atoms.

(f) *Release of charged carriers from surfaces and electrodes*

Electrons, positive ions and negative ions may be released from electrodes or from the walls of a discharge tube. The electrode emission is, of course, fundamental to the operation of the discharge, but any wall emission is merely a by-product of the discharge processes and, in fact, is usually quite unimportant. Emission processes at the anode, although these do apparently take place, are also of relatively little importance.

The emission of electrons from the cathode may be caused by bombardment by positive ions, by photoelectric emission, by thermionic or field emission or by metastable atoms. Secondary emission due to electron bombardment does not take place, on account of the fields repelling electrons from the cathode.

As individual processes, thermionic emission and field emission are now fairly well understood, and nothing of importance to gas-discharge phenomena has appeared in recent years. Less is known about the effect of positive-ion bombardment, and there are considerable experimental difficulties in obtaining accurate data directly. The well known complications due to the presence of gas films on the surface result in many of the measurements obtained being applicable only to the particular conditions of the surface studied at the time. There does not yet appear to be a satisfactory wave-mechanical theory of the process. It seems that all high-energy positive ions (>100 volts) liberate electrons from metal surfaces, and, provided the ionization potential of the atom concerned is greater than the work function of the surface, ions with much lower energy will be effective (Penning, 1928). Data on the yields at various energies for various reproducible surfaces are practically non-existent; indirect evidence, however, makes it fairly certain that positive ions of the low energies normally occurring in discharges are, to some extent, effective in releasing electrons. So far as photo-electric emission is concerned, the ground has been well explored at wave-lengths above 2000 A. Much less is known for the region below 1000 A., which may be of great importance (see, for example, Kenty, 1933). That metastable atoms may be effective in releasing electrons from surfaces, provided their energy is greater than the work function of the surface, is known, but it is doubtful if this process is ever of very great importance in discharge-tube phenomena (Kenty, 1933).

In discharge devices there are two principal types of electrode employed. Negative glow lamps and the high-voltage signs used for advertising employ relatively large and usually unactivated metal surfaces which take currents of the order of 100 ma., with current densities of some 1 to 100 ma. per cm^2 of electrode surface. The high-pressure Hg vapour lamps and Na lamps used for illumination purposes employ a smaller activated electrode which takes currents of several amperes, and the current density may be hundreds of amperes per cm^2. In the former type, positive-ion bombardment and, to a lesser extent, photoelectric emission, are the effective agents in releasing electrons; and in the latter type thermionic emission is predominant. In both cases, when the discharge is initiated, the electrodes are frequently cold, and the starting is then unassisted by thermionic emission. The emission mechanism operative at the mercury-pool cathodes of rectifiers is not fully understood ; whether it is field emission or thermionic emission, or even whether these terms retain any meaning in the peculiar conditions obtaining at the surface, is not clear (Loeb, 1939 a).

There are some recent papers of interest dealing with emission from solid bodies. Tonks (1938) discusses the conditions existing at the surface of a mercury-pool cathode. Guntherschulze *et al.* (1938 a, 1938 b) have measured indirectly the yield of electrons by positive-ion impact at cathodes of various metals (K, Cu, Ag, Fe) in He, Ne and H_2 glow discharges. In all cases, the number of electrons released per positive ion (γ) varied monotonically from some 0·2 at a cathode fall of 500 volts to about 1 at 2500 volts. Kruithof and Penning (1936) and Jones and Willott (1936 b) have made similar measurements for Cu cathode in an A glow discharge. Kruithof and Penning used a range of X/p (X is the field strength in volts per cm., and p the gas pressure in mm. Hg) of from 5 to 400 (voltage gradient from 100 volts to 2000 volts and pressures from 0·7 mm. to 150 mm.). Jones and Willott's experiments were carried out over a similar range of X/p ; γ, with a value of 0·1 to 0·2, seemed substantially independent of X/p. Townsend and Yarnold (1934) with an Al cathode in a He discharge found a linear dependence of γ on X/p, varying from 0·015 at $X/p = 30$ to $\gamma = 0{\cdot}055$ at $X/p = 70$.

An interesting paper by Paeton (1939) describes electron emission from anode and cathode after the discharge is switched off. This is explained as being due to the action of photons in the discharge. Arnot and Beckett (1938) have described a new process, whereby negative ions are formed by bombarding metallic surfaces with positive ions. The process appears complex, as it seems that an excited neutral atom is formed as an intermediate product. Smith (1938) has studied the question theoretically. This process is of great interest, but there is little evidence that it is important from the point of view of discharges in gases.

(g) *Recombination*

In gases in which the negative ion comprises one or more atoms, recombination is not really an individual process, since the probability of recombination of two ions of opposite sign may depend more on the distribution of other ions than on their own position and velocity. It is true that ultimately the two ions

may come into "wave-mechanical contact", and at this stage the collision may be truly considered an individual process. Very little is known about this type of interaction, however, and data on the parameters of the various types of collision are not available. Ionic recombination is not of very great importance in most discharge tubes, but its study has recently become of practical value in connection with the detection of cosmic radiation by means of the high-pressure ionization chamber (see, for example, Zanstra, 1935; Kraus, 1937; Broxon *et al.*, 1938 a). The most notable advance in the last ten years or so has been in the clarification of the understanding of the various types of recombination which may take place. This is dealt with in detail by Loeb (1939 a). There are the "volume recombination", which occurs when ions are randomly distributed in space, and to which Thomson's diffusion theory (1924) applies, provided the pressure is not much above one atmosphere; the "preferential recombination", when the electron forms the negative ion so close to the parent positive that the electrostatic attraction is predominant, and the same ions ultimately recombine: this occurs at higher pressures (3 atmospheres and upwards in air), and in this case Langevin's theory (1903 a, 1903 b) becomes applicable; "initial recombination", which occurs before diffusion processes are able to lead to a random distribution; and "columnar ionization", which occurs in densely ionized tracts, for example those left by α-particles. Each of these types of recombination has its own coefficient of recombination which is different in value from the others, and may vary with factors such as pressure in a different way. It is now clearly realized that all the early measurements were more or less meaningless. The different types of recombination were inextricably confused; moreover, the results were complicated by the difficulties inherent in the technique of measurement. The interpretation of recent experiments, however (Gardner, 1938; Sayers, 1938; Machler, 1936 a, 1936 b; Onsager, 1938), gives good hope that the mechanism is now fairly well understood.

In gases in which negative ions do not form, recombination may take place either by means of electron, positive-ion recombination in the gas or by diffusion of electrons to the walls, where the probability of recombination is very high. These processes are of much greater importance in most discharges than ionic recombination. At low pressures, where the electron mean-free-path is not very small compared with the diameter of the tube, and the electron velocities are large, most of the electrons drift to the walls before recombining. At higher pressures, however, where the mean free-path becomes small and the electron velocities are also small, appreciable recombination in the body of the gas may take place, and this is particularly so when the current density and the electron and positive-ion densities are high. The probability of the recombination of electrons and positive ions is, therefore, of great importance. The electron may combine into one of the many levels in the atom, and for each level there is a value of the effective cross-section for recombination, which is a function also of the relative velocity of electron and ion. Very few measurements on these cross-

sections have been made, although it is known that they are small, probably of the order 10^{-20} cm^2.

Mohler *et al.* (1929, 1933 b) have used measurements on the intensity distribution in the recombination spectrum to estimate the cross-section for recombination into the various levels in the case of Cs vapour. The intensity is given by

$$\frac{\mathcal{J}(\nu)}{h\nu}d\nu = \Sigma N^-(v) \,.\, N^+ \,.\, v \,.\, q(v\nu_1)\, dv, \qquad \ldots\ldots(1)$$

where the summation extends over all limits of frequency (ν_1) less than ν and where

$\mathcal{J}(\nu)$ is the intensity at any frequency $\nu(>\nu_1)$;

$N^-(v)$ is the concentration of electrons with velocity v;

N^+ is the concentration of positive ions;

$q(v\nu_1)$ is the cross-section for recombination of an electron with velocity v into the level ν_1.

Thus, if $N^-(v)$ and N^+ can be found (for example, by probe methods), $q(v\nu_1)$ can be calculated from $\mathcal{J}(\nu)$. They found the effective cross-section to be of the order of 10^{-19} cm^2 for recombination into the $2P$ level of Cs at an electron energy of 0·2 volts, and 10^{-21} cm^2 for 0·3-volt electrons into the $6P$ level. As the electron energy decreased, the cross-section increased rapidly. The results were in fair agreement with the theoretical relation:

$$vq(v\nu_1) = \frac{K}{\nu^3(\nu-\nu_1)^{\frac{1}{2}}}, \qquad \ldots\ldots(2)$$

where K is a constant.

From the point of view of discharge mechanism, the more important data are the total probabilities for the recombination of an electron and ion under various conditions, that is, the sum of the probabilities of recombination into the various levels.

Theoretically, the above method would give the total cross-section if all the cross-sections for recombination into the various levels as a function of electron velocity could be found. From the total cross-section for recombination, the recombination coefficient could be calculated as a function of the electron-velocity distribution provided there were no other influences acting. It is not convenient in practice to determine the total cross-section by this summing process, since the intensity measurements for all the levels concerned present great difficulties. Moreover, even if the total cross-section for recombination could be obtained in this way, the measurements would be only of limited application. The reason for this is that the probability of recombination may be considerably influenced by the presence of other atoms, so that the values obtained would probably be valid only at the pressures at which the experiments were carried out, and could not be applied as true atomic parameters.

The actual recombination coefficient will therefore be a function of gas pressure as a result of the influence of neighbouring atoms, as well as being a function of the

electron-velocity distribution. A few measurements have been made on recombination coefficients, but there are no data available to show the dependence on electron temperature and gas pressure. One of the methods used is to measure the intensity in the continuous spectrum in the afterglow at various times after stopping the discharge. Kenty (1928) made measurements on A in this way and Mohler (1937 a) on Cs vapour. If the light emitted is due to recombination, the intensity j is proportional to n_e^2, where n_e is the electron concentration. It follows, therefore, that

$$j_2^{-\frac{1}{2}} - j_1^{-\frac{1}{2}} = k\alpha(t_2 - t_1), \qquad \ldots\ldots(3)$$

where j_2 and j_1 are the intensities measured at intervals t_2 and t_1 after the discharge is stopped, and k is a constant. Any loss of electrons other than by recombination will show in the deviation from linearity of a plot of $j^{-\frac{1}{2}}$ *vs.* t. The constant k is obtained by assuming that the intercept on the axis of $j^{-\frac{1}{2}}$ at $t=0$ is proportional to n_{e0}, the electron concentration in the steady discharge, which can be obtained by probe methods. The value of α obtained in Cs in the pressure range 46 to 110 μ, and with an electron temperature of 1200° K., was about $3{\cdot}5 \times 10^{-10}$. The value of α will, if measured by this method, depend on the rate of change of electron temperature after the discharge is stopped, and allowance should be made for this. Mohler (1937 b) has also used probe methods to study recombination in the Hg positive column. Again a figure of the order of 10^{-10} was obtained for a pressure of 270 μ and an electron temperature of 2000° K. To what extent this value can be considered the true recombination coefficient, or to what extent it is influenced by the presence of other atoms, is not clear. The order of the coefficient for electron recombination should be compared with the figure of 10^{-6} for ionic recombination.

§ 3. AVERAGE AND MACROSCOPIC PHENOMENA

(a) *Electron velocity distribution*

The primary ionization process by electron collision is fundamental in any discharge, and as excitation almost invariably accompanies ionization, the two processes have usually to be considered together. It is clear that, for a given mean electron energy, the mechanism of excitation and ionization, and thus the energy balance of the discharge, depend largely on the distribution of electron velocities. A knowledge of the distribution of electron velocities as well as of the mean electron temperature, therefore, is necessary in almost any attempt to interpret discharge phenomena.

There is no satisfactory way of determining the distribution experimentally (see, however, section (d), on Probes), and a clear insight into the way in which the various types of collision and the electron density affect the velocity distribution has, until recently, been lacking. The velocity distribution results from the statistical equilibrium between the energies gained in the field and lost at collisions. It is easy to see, in general terms, how this comes about when it is remembered that all the processes involved are functions of the electron velocity. Suppose,

for example, the velocity distribution is $\rho(v)$; if all the necessary atomic parameters were known as functions of the electron velocity, it would be possible formally to calculate the distribution of mean free-paths for any particular type of collision as a function of $\rho(v)$. From this, the gain of energy between collisions and the energy lost at collisions could be calculated, and the two being equated should give $\rho(v)$. The problem in the general form is very complicated, and the distribution obtained theoretically depends upon the simplifying assumptions which have to be made. Nevertheless, within the last ten years theory has gone a long way towards the complete solution.

When the velocity distribution is known, the mean electron energy is given by

$$\bar{\epsilon} = \int_0^\infty \epsilon\rho(\epsilon)d\epsilon \Big/ \int_0^\infty \rho(\epsilon)d\epsilon, \qquad \ldots\ldots(4)$$

where ϵ is the energy of an electron.

For the study of the electron velocity distribution, it is desirable that the discharge conditions should be as simple as possible in order that the share of each individual process concerned in promoting the distribution should be understood. For this reason, the positive column is particularly suitable for the study. Here the field is uniform along the axis, so that the velocity distribution is sensibly constant along the axis. All the early probe measurements agreed in predicting in the positive column a distribution that was approximately Maxwellian under the conditions in which the probe theory was valid, but it was known that it was unlikely that this would be universally true, and, moreover, it was not clear how the Maxwellian distribution was produced.

Druyvesteyn (1930 b) was the first to attack the problem theoretically in a restricted form. He assumed that inelastic collisions could be neglected, that there was no interaction between electrons, that the mean energy of the atoms was small compared with that of the electrons, and that the electron mean-free-path was independent of the electron velocity. It is obvious that these assumptions can never be accurately true in a discharge which is self-maintained, since some ionization is always necessary, and therefore inelastic collisions must take place. Nevertheless, the solution of the problem is important as an idealized case. The distribution function obtained is

$$\rho(v) = Cv^3 \exp\left[-\frac{3\alpha}{8\lambda_e^2\gamma^2}v^4\right], \qquad \ldots\ldots(5)$$

where λ_e is the mean free-path, $\gamma = X_e/m$, m is the mass of the electron. $\alpha = 2m/M$, M is the mass of the atom or molecule, and C is a constant. This differs from the Maxwellian distribution essentially in having the fourth power of the velocity in the exponential. Examples of the Maxwellian and Druyvesteyn distributions are shown in figure 5 for the same mean energy. At low values of the field strength, where the electrons are more nearly in equilibrium with the gas molecules, the distribution would be expected to be nearly Maxwellian. The Druyvesteyn distribution reflects the effect of the field in disturbing this distribution.

Morse, Allis and Lamar (1935), using the same assumptions of elastic impact and no electron interaction, dealt with the problem in a rather more general manner. The method used was a development of one originated by Lorentz in his work on electrons in metals, in which the numbers of electrons entering and leaving an element of volume in phase space are equated. Two equations are obtained which correspond to the detailed balancing of energy and momentum. These give the velocity distribution in the particular case of homogeneous steady conditions in the form

$$p(v) = A \exp\left[-\frac{3m^2}{M} \int_0^v \left(\frac{NQ}{eX}\right)^2 v^3 \, dv \right], \qquad \ldots\ldots(6)$$

where Q is the momentum transfer cross-section (see section (b), on Mobility of electrons) and A is a constant. If the cross-section is a function of the electron velocity, the equation allows the effect of this variation to be seen. In the event

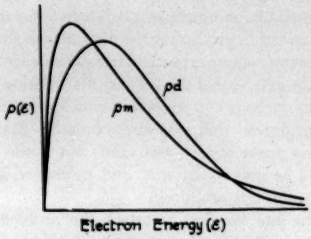

Figure 5. Maxwellian distribution (p_m) and Druyvesteyn distribution (p_d) for the same electron mean energy.

of the Ramsauer effect being present, the distortion of the distribution on this account may be considerable (Allen, 1937). When the mean free-path is constant, the distribution reduces to that of Druyvesteyn.

Davydov (1935) derived the same formula with the same assumptions as Druyvesteyn, except that he extended his analysis to take account of the thermal agitation of the atoms, and showed thereby that when the fields were small the distribution was Maxwellian.

The assumption that inelastic collisions may be neglected restricts the above formulae to the case where the fields are large enough to produce a mean energy of agitation of the electrons large compared with that of the atoms, but not large enough to cause appreciable excitation. When this restriction is removed, the problem becomes much more complicated. Another important factor which the treatments mentioned above do not take into account is the interaction between electrons.

Druyvesteyn (1936) made the first attempt to consider the effects of inelastic collisions. He calculated the velocity distribution for a Ne discharge assuming

only one ionization and one excitation potential. Smit (1937) has, however, made the most comprehensive analysis to date and has applied his formulae to calculate the velocity distribution in the case of He for values of X/p from 3 to 10. He takes into account elastic collisions, exciting and ionizing collisions, the mean energy of the atoms and the variation of mean free-path with velocity, but assumes symmetrical scattering. The formulae of Maxwell, Druyvesteyn and Davydov are obtained as special cases of the general equation. The analysis is, however, lengthy and complex, and the results cannot be expressed in a simple form in general terms. Smit's curves for the velocity distribution in He are shown in figure 6.

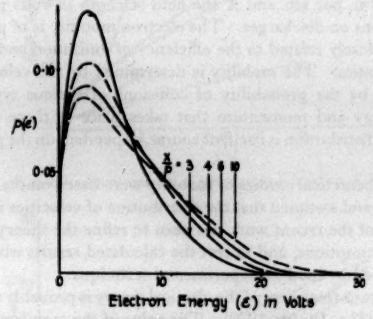

Figure 6. Calculated electron-velocity distribution in He for various values of X/p (Smit).

When ionizing collisions are included, no steady state can be reached unless some mechanism for the removal of electrons is postulated, for unless this is done, the density of electrons will increase in the direction of the field. With this understanding, however, that the electron density is no longer independent of position, the distribution of velocities may still be found. In a practical case, as, for instance, the positive column, the distribution of velocities itself will naturally depend upon the method of removal of the electrons. This consideration introduces a further considerable complication into the analysis, and so far the solution of the problem does not seem to have been attempted.

Smit examined the effect of electron interaction. At densities less than about 10^{-3} of that of the atoms the effect is small; above this value the interaction helps to bring about the Maxwellian distribution. This question has been considered also by Gabor (1932, 1933 a, 1933 b), Davydov (1936 b, 1937) and Haseltine (1939). Davydov considers that in the positive column the radial field may have an important effect in bringing about the Maxwellian distribution. There seems to be a real doubt whether the electron interaction is sufficient to account for the Maxwellian distribution in all cases where it is known to exist.

With the exception of this question, which is still somewhat obscure, the theoretical work of the last few years has led to a clear understanding of the electron energy distribution. When accurate data concerning the collision processes are available, the methods mentioned above allow in certain circumstances exact quantitative calculation of the electron-velocity distribution to be made. This is a most important step towards bridging the gap between the individual and the average processes in gas-discharge theory.

(b) *Mobility of electrons*

The value of the mobility of charged particles (i.e. $\mu = u/X$, where u is the mean drift-velocity in cm. per sec. and X the field strength in volts per cm.) appears in most calculations on discharges. The electron mobility is of particular importance since it is closely related to the efficiency of ionization and to other fundamental characteristics. The mobility is determined by the velocity distribution of the electrons, by the probability of collisions of various types, and by the exchange of energy and momentum that takes place at these collisions. The electron-velocity distribution is itself, of course, dependent on the primary collision processes.

All the older theoretical studies of mobility were based on the classical elastic-collision concept, and assumed that the distribution of velocities was Maxwellian. The aim of most of the recent work has been to refine the theory by introducing more accurate assumptions, and to test the calculated results with measurement of mobility obtained by modern experimental technique.

The most rigorous treatment of the classical theory is probably that of Compton and Langmuir (1923 a, 1923 b, 1930). The value of the mobility is given by

$$\mu_e = \frac{0{\cdot}921 e\lambda_0}{(2em)^{\frac{1}{2}}}\left[\frac{\Omega}{2} + \left(\frac{\Omega^2}{4} + \frac{\lambda_0^2 X^2 M}{5{\cdot}32m}\right)^{\frac{1}{2}}\right]^{-\frac{1}{2}}, \qquad \ldots\ldots(7)$$

where $\lambda_0 = (\pi\sigma^2 n)^{-1}$, $\pi\sigma^2$ is the collision cross-section, n is the density of the molecules, M is the mass of a molecule and m that of an electron, whilst $e\Omega$ is the mean energy of agitation of the molecules. In the derivation it is assumed that both electrons and molecules have Maxwellian distributions, that the collisions are elastic and that the mean free-path is independent of the electron velocity.

In the case of large fields, equation (7) reduces to :

$$\mu_e = \left(\frac{\lambda_0 e}{Xm}\right)^{\frac{1}{2}}\left(\frac{m}{M}\right)^{\frac{1}{2}}, \qquad \ldots\ldots(8)$$

and in the case of small fields to

$$\mu_e = \frac{0{\cdot}921 e\lambda_0}{(2em\Omega)^{\frac{1}{2}}}. \qquad \ldots\ldots(9)$$

Since for small fields $e\Omega$ becomes equal to the equipartition energy $\frac{1}{2}mC^2$, where C is the r.m.s. velocity of the electrons, (9) may be written in the more usual form

$$\mu_e = \frac{0{\cdot}921 e\lambda_0}{mC}. \qquad \ldots\ldots(10)$$

Thus, if only elastic collisions were operative and the other assumptions applied, the mobility at small fields would be independent of field strengths and at large fields would be inversely proportional to the square root of the field strength. The more general methods, for example that of Morse, Allis and Lamar (1935),

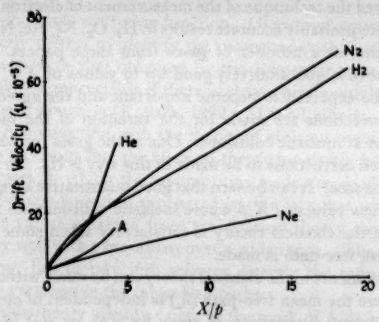

Figure 7. Drift velocity of electrons as a function of X/p.

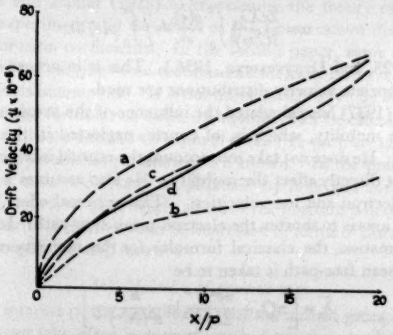

Figure 8. Drift velocity of electrons in H_2.

a. Calculated from equation (7).
b. Allowance made for variation of mean free-path with velocity.
c. Allowance made for inelastic collisions.
d. Experimental curve.

give the same results (with the possible exception of small differences in constants) when the same assumptions are made.

The agreement with experiment has shown that equation (7) gives the correct trend of μ_e with varying X, and the order of magnitude of μ_e is also correctly

given. There is little doubt, therefore, that the theory is basically satisfactory and that the errors almost certainly lie in the crude assumptions made regarding the collision processes.

On the experimental side Nielsen and Bradbury (1936 a, 1936 b, 1937) have recently improved the technique of the measurement of electron mobility and have obtained what are probably accurate results in H_2, O_2, N_2, He, Ne and A. Figure 7 shows the results for a number of gases from these papers. Agreement with equation (7) above is quantitatively good up to values of X/p, at which inelastic collisions may be expected to become important and the agreement is improved when rough corrections are made for the variation of the mean free-path and the energy losses at inelastic collisions. One of the gases for which data are available to allow such corrections to be made in this way is H_2. In figure 8 are shown the results in this case. It can be seen that good quantitative agreement is obtained. In all cases, at low values of X/p where inelastic collisions do not take place, the agreement with the classical theory is satisfactory when some allowance for the variation of mean free-path is made.

Several refinements to the classical theory can be made without great difficulty. In the case where the mean free-path (λ_ϵ) is independent of electron velocity and the electron velocity distribution is given by $\rho(v)$, or $\rho(\epsilon)$ in terms of electron energies, the mean velocity of drift is given by

$$u = \frac{2eX\lambda_\epsilon}{3(2m)^{\frac{1}{2}}} \int_0^\infty \frac{\rho(\epsilon)}{\epsilon^{\frac{1}{2}}}\, d\epsilon \Big/ \int_0^\infty \rho(\epsilon)\, d\epsilon. \qquad \ldots\ldots(11)$$

(See Hertz, 1925, and Druyvesteyn, 1936.) This is in accord with (8) and (9), when the appropriate velocity distributions are used.

Gvosdover (1937) has calculated the influence of the presence of positive ions on the electron mobility, which is, of course, neglected in the classical elastic-collision theory. He does not take into account the mutual interaction of electrons, as this does not directly affect the mobility. He also assumes a Maxwellian distribution for electron and ion velocities. The principal effect of the presence of the positive ions is to shorten the electron mean-free-path. It is found that to a first approximation, the classical formulae for the mobility may still be used, provided the mean free-path is taken to be

$$\bar{\lambda}_\epsilon = \left[nQ_1 + \frac{\pi e^4 N^+}{2k^2T^2} \ln \frac{kT}{e^2(N^+)^{\frac{1}{3}}} \right]^{-1}, \qquad \ldots\ldots(12)$$

where n is the concentration of neutral molecules, N^+ is the concentration of positive ions, and Q_1 is the total cross-section for electron collision with the neutral molecules.

In Gvosdover's work, as in that of others who use the method of equating the momentum gained by an electron in the field to that lost at collisions, the cross-section for momentum transfer appears. For example, this is given for positive ions by

$$Q^+(v) = 2\pi \int_0^\pi \sigma^+(v, \theta)(1 - \cos\theta) \sin\theta\, d\theta \qquad \ldots\ldots(13)$$

whereas the cross-section for absorption on which the mean free-path in the Ramsauer type of experiment depends is

$$q^+(v) = 2\pi \int_0^\pi \sigma^+(v, \theta) \sin\theta \, d\theta, \qquad \ldots\ldots(14)$$

where $\sigma^+(v, \theta)$ is the angle scattering function. Gvosdover points out, therefore, that it is not strictly correct to identify the mean free-path used in momentum-transfer calculations with that obtained by the Ramsauer type of experiment. The difference may be appreciable if there is strong preferential scattering in the forward or backward direction (Kollath, 1932; Maier-Leibnitz, 1935).

The position in connection with the theory of electron mobility is thus satisfactory. There is no doubt that the classical theory for elastic spheres is approximately valid in cases in which the assumptions made are good approximations to the truth. This conclusion is supported by the fact that when approximate corrections to the values obtained from the classical theory can be applied, the results agree fairly well with experiment over a wide range. The further development of the theory has to take into account directly the influence of inelastic collisions and the variation of the mean free-paths.

The basis of this theory already exists in the work of Allis and Allen (1937), Allen (1937) and Smit (1937). The first-named authors have applied the method of Morse, Allis and Lamar (1935) to re-examine the theory of the Townsend diffusion-drift experiment and the effect of the Druyvesteyn distribution on the mobility and diffusion coefficients. In the second paper, some attempt is made to study the effect of varying cross-sections and inelastic impacts on the mobility, but the treatment is incomplete. It is not immediately obvious, but it seems probable that, from the formulae of Smit, the mobility can be calculated for the case where inelastic collisions and the variation of mean free-path are taken into account when the necessary collision parameters are known. In the present form of the theory, however, results can only be obtained by long and involved numerical calculations. Even so, to have nearly attained the position where so complicated a process as the drift of electrons can be expressed in terms of the atomic processes must be considered a big step forward.

(c) *Mobility of positive ions*

The recent interest in the mobility of positive ions in gases in which cluster formation does not take place is derived partly from the possibilities it provides of studying the ion-molecular forces, although in fact it appears that it is rarely possible to obtain accurate atomic data from this type of work. The theoretical formulae for the mobility depend upon the assumptions made with regard to the intermolecular forces. The extent of the agreement between experimental observations and calculated values of the mobility, and of the variation of mobility with other factors such as temperature and pressure, may give information about the validity of the assumptions. It is true that in the derivation other assumptions may also be made (e.g. that the distribution of velocities is Maxwellian), but for positive ions the degree of validity of these assumptions and the effect on the results of departures from them can usually be estimated.

The simple theory of Langevin (1903 a, 1903 b, 1905) gives for the mobility

$$\mu_+ = \frac{0{\cdot}815e\lambda_0}{M_1 C_1}\left(1+\frac{M_1}{M_2}\right)^{\frac{1}{2}}, \qquad \ldots\ldots(15)$$

where $\lambda_0=(\pi n\sigma_{12}^2)^{-1}$, $\pi\sigma_{12}^2$ is the collision cross-section, M_1 is the mass of the molecule, M_2 is the mass of the ion, and C_1 is the r.m.s. velocity of the molecules.

It is assumed in this that the ions and molecules are in temperature equilibrium with a Maxwellian velocity-distribution, that the particles may be treated as solid elastic spheres, that there are no forces between the charged and uncharged particles, and that the ion drift-velocity, which is superimposed on the Maxwellian velocity-distribution, is small compared with the mean energy of agitation. The equation is therefore applicable only in cases of small X/p. Compton and Langmuir (1930) have modified the formula to take account of a difference in mean energy between the atoms and ions, and thus derived a formula which should be applicable at higher values of X/p. If the ions are singly charged, they find

$$\mu_+ = \frac{0{\cdot}921e\lambda_0}{(2eM_2)^{\frac{1}{2}}}\left[\frac{\Omega}{2}+\left\{\frac{\Omega^2}{4}+\frac{\lambda_0^2X^2(M_1+M_2)^2}{5{\cdot}32M_1M_2}\right\}^{\frac{1}{2}}\right]^{-\frac{1}{2}} \times\left[1+\frac{M_2}{M_1}\frac{1}{\frac{1}{2}+\left\{\frac{1}{4}+\frac{\lambda_0^2X^2(M_1+M_2)^2}{5{\cdot}32M_1M_2\Omega^2}\right\}^{\frac{1}{2}}}\right]^{\frac{1}{2}}, \qquad \ldots\ldots(16)$$

where $e\Omega$ is the mean energy of agitation of the gas molecules. If, as is usually the case in the important applications of discharges in gases, there is no cluster formation, $M_1=M_2$.

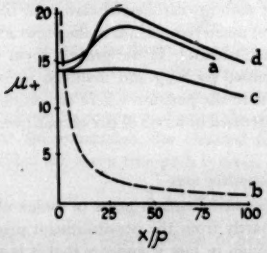

Figure 9 *a*. The mobility of K^+ ions in H_2.

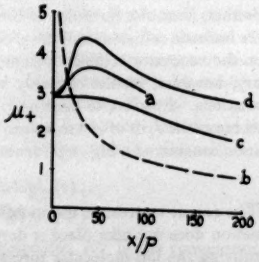

Figure 9 *b*. The mobility of K^+ ions in A.

Curve *a*. Measured values.
Curve *b*. Calculated from equation (16).
Curve *c*. Calculated from Hershey's formulae for the extension of Langevin's theory.
Curve *d*. Calculated from Hershey's formulae for the extension of Hassé and Cook's theory.

Hershey (1939 a) has recently measured with improved technique the mobility of positive ions in H_2, He, Ne and A up to values of X/p of about 100. Figures 9 *a* and 9 *b* (curves *a*) show the measured values of the mobility (μ_+)

for K^+ ions in H_2 and A respectively. Curves *b* show the values calculated from equation (16). The agreement of the measured values with those calculated from equation (16) is not good, although it is better in A than in the lighter gas. This would be expected, since the ion random-energy in the latter case is not large compared with the drift energy as assumed in the derivation. Except at very low pressures, the effect of the ion-molecular forces is probably appreciable.

Equations (15) and (16) fail in that they do not give the correct absolute values for the mobility nor the correct variation with other variables such as temperature and X/p. The reason for this must be sought in the assumptions made with regard to the collision processes. As long ago as 1905, Langevin developed a theory which took into account the effect of the ion-molecular forces. Although the theory was produced so many years ago, it is only comparatively recently that its value has been realized. He considered the molecules as elastic spheres, the uncharged molecules being polarizable in the field of the ions, and the attractive force being given by $\frac{(\epsilon-1)e^2}{2\pi r^5}$, where ϵ is the dielectric constant of the gas and r is the distance between the spheres. The expression obtained for the mobility is

$$\mu_+ = \frac{f(\beta)}{\sqrt{\rho(\epsilon-1)}}\left(1+\frac{M_1}{M_2}\right)^{\frac{1}{2}}, \qquad \ldots\ldots(17)$$

where $f(\beta)$ is a known function of β and $\beta^2 = \frac{8\pi p\sigma_{12}^4}{(\epsilon-1)e^2}$, p is the pressure of the gas, and ρ is its density.

In this equation, the only uncertain quantity is σ_{12}, which is not defined for actual molecules. Hassé and Cook (1929, 1931) have attempted to improve on this by considering the particles as point centres of force; in this case the collision cross-section depends only on the relative velocities of the particles and the law of force. They consider an inverse-fifth power law of attraction superimposed upon the dielectric forces together with an inverse-ninth power law of repulsion, the resultant attractive force being

$$\frac{\mu_a}{r^5} - \frac{\lambda_r}{r^9}, \qquad \ldots\ldots(18)$$

where r is the distance between ion and molecule, and μ_a and λ_r are atomic constants. The expression obtained for the mobility, while complicated and different analytically from equation (17), is similar in the results which it gives. It contains a known but complicated function of $s = \frac{2RT\lambda_r}{\mu^2 a}$ in place of $f(\beta)$, and the cross-section σ_{12} in equation (17) is thus replaced by the constants μ_a and λ_r. Some information about these constants may be obtained from viscosity measurements (Hassé, 1929) and, where this information is available, the theory is more fundamental than that of Langevin. The powers of the attractive and repulsive

forces, while comparatively simple from the point of view of the rather complicated analysis, are probably not those which best fit the facts. Inherent in both Langevin's and Hassé and Cook's derivations are the assumptions that the drift velocity is small compared with the velocity of thermal agitation, that is, the theory is valid only for weak fields. It is assumed also that there is no cluster formation, and the possibility of electron transfer is not taken into account.

Making the best possible estimates for σ_{12} in Langevin's equation, or for μ_a and λ_r in Hassé and Cook's equation, the calculated values agree with the observations for weak fields at least as well as can be expected, and these derivations represent a marked advance on the simple theory as represented by equation (16). It would seem that the fairly satisfactory agreement between theory and experiment here should settle the old controversy in connection with cluster formation, since there is no need to postulate cluster formation at least for young ions (10^{-4} sec.) in H_2 and N_2 or in the rare gases.

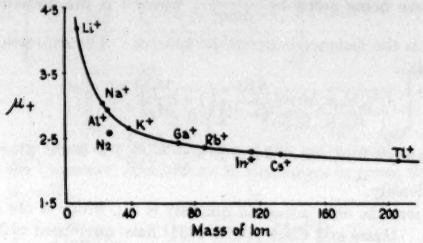

Figure 10. The mobility of positive ions in N_2.

In addition to the test of the absolute values given by the above theories, other tests can also be applied: the variation with mass of ion in a given gas and the temperature variation. Langevin, Hassé and Cook and the classical theory all agree on the mass factor. The mobilities of a number of ions in N_2 have been measured (Powell *et al.*, 1932; Brata, 1933; Mitchell *et al.*, 1934), and it is seen from figure 10 that, with the exception of the N_2 ion itself, all the points lie on a smooth curve:

$$\mu_+ = A\left(1+\frac{M_1}{M_2}\right)^{\frac{1}{2}} \qquad \ldots\ldots(19)$$

(where A is a constant) connecting the mobility with the mass of the ion. The exception in the case of N_2 is probably explainable by the effect of electron exchange, which should lead to a smaller value for the mobility (Hassé *et al.*, 1931). This observed variation with mass of ion should occur according to the theory, provided the variation of the " size " of the ion is small. There is a range ($p<1$) within which the coefficient of the mass factor in equation (17) is insensitive to

changes in β, and this is also the case in Hassé and Cook's expression, although the physical reason for this does not seem to be clear. The measurements concerned are probably all within this range, in which case the "size" effect would be small. Tyndall and Powell (1932) measured the mobility of alkali ions in A, Ne and He. For A, the points fitted well with the curve of equation (19); they did not fit so well for Ne, and were still worse for He. These results may quite well be expected from variations in the "size" of the molecules. The variation with ion mass seems, therefore, to agree with theory.

According to both Langevin's and Hassé and Cook's theory, the mobility at constant density is a complicated function of the temperature, which should have a maximum value. The experimental results (Tyndall *et al.*, 1935; Pearce, 1936; Huxley, 1937) show a fair agreement with the theory. There is some evidence from these experiments that agreement with Hassé and Cook's formula would be improved if a higher inverse-power than 9 were used for the repulsive force.

It seems, therefore, that at small values of X/p, where the assumptions made in Langevin's and Hassé and Cook's formulae are valid, the situation is as satisfactory as can be expected in view of the scarcity of data on the ionic forces. At larger values of X/p the only theory until recently was that of Compton (equation (16)). Hershey (1939 b) has, however, extended the results of Langevin and Hassé and Cook to fields of any strength. It is assumed that there are no inelastic impacts and that the random energy of the ions has a Maxwellian distribution. The drift and random energies are evaluated by equations of momentum and energy balance. Curves *c* and *d* in figures 9 *a* and 9 *b* show the results obtained from the Langevin and Hassé and Cook models respectively. It can be seen that the qualitative agreement with experiment (Hershey, 1939 a) is good; in particular the peaks observed in the {mobility *vs.* X/p} curves are accounted for.

So far as the mobility of ions in their own gas is concerned, it would seem that the greatest factor of uncertainty at the moment is that of electron exchange, which has the effect of decreasing the mobility. Further data on this point are needed before any of the above theories can be applied with confidence to this important case.

(d) *Probes*

The probe methods originated by Langmuir and Mott-Smith (1923 and 1923 b, 1925) are still of great value in the study of the conditions in a discharge. The use of probes allows much useful information to be obtained with reasonable certainty. It is still the only way of obtaining direct experimental data on the velocity distribution of electrons and ions, although, in view of the many doubtful factors involved, it is somewhat unsatisfactory. The interpretation of probe measurements is simplest when the electron and positive-ion concentration are high, and when the fields are low—that is, in conditions such as exist in the plasma of the positive column. The original probe theory made a number of assumptions which are discussed in Langmuir's original papers (see also Loeb, 1939 a). In

circumstances in which these assumptions are valid, the probe current *vs.* voltage curves, rightly interpreted, give the electron temperature and concentration, the positive-ion temperature and concentration, the space potential and the wall (or probe) potential.

The definite limitations of the method arise from the limited range of conditions in which the simple probe theory is valid. In practice it is not always possible to know when these assumptions are satisfied, nor easy to estimate what effect deviations from them will have on the interpretation of the results of experiments. One of the major assumptions made is that the electron-velocity distribution is Maxwellian. When this is the case, the interpretation of the curves is particularly simple; when it is not the case, but, nevertheless, the velocity distribution is known, some progress may be made in deriving data on the conditions in the discharge. When the electron-velocity distribution is not known, probe methods could not until recently be used at all.

The most important development of the technique of probe measurements in recent years is in this connection. Druyvesteyn (1930 a) extended the theory of probes to cases in which the electron-velocity distribution was not Maxwellian. He showed that the second derivative of the electron probe-current (i_-) in the negative probe-voltage region with respect to the probe potential (v_s) could be utilized to give the electron-velocity distribution $p(v)$ by means of an equation of the form

$$p\left(\sqrt{\frac{2eV}{m}}\right) = \frac{4m}{Ae^2} V \frac{d^2 i_-}{dV_s^2}, \qquad \ldots\ldots(20)$$

where A is the area of the probe and V is the difference between the space potential and probe potential.

Further progress has been made on these lines by Sloane and McGregor (1934), who developed an alternating-current valve method which allowed the second derivative to be obtained directly without the necessity for double integration, and by Van Gorcum (1936), who used a bridge circuit and a cathode-ray oscillograph to measure the second derivative. The latter author studied the velocity distribution in a Ne discharge and obtained very interesting results, showing how, as the positive column was approached from the cathode side, the velocity distribution gradually changed into the Maxwellian.

Other attempts to extend the theory and field of application of probe methods are those of Davydov and Zmanovskaja (1936 a) on the generalization of the theory for probes of any size; and of Borngis (1938), who investigated the effect (neglected in the original theory) of collisions within the probe sheath at high pressures. A paper by Gorschkow and Maslakowery (1936) describes an interesting possibility which might be capable of development, namely, the use of a thermo-probe consisting of a copper-constantan thermocouple.

There have been some attempts to apply probe methods to arcs at high pressures when the gas itself is at a high temperature. Difficulties then occur due to overheating of the probe and possible thermionic emission. Nottingham

(1928, 1929) and Bramhall (1931, 1932) have tried moving the probe rapidly to avoid these complications. Mason (1937) criticizes the method and shows that the cooling effect of the probe can have a serious disturbing effect. There seems no doubt that, in the present state of technique, probe methods cannot give quantitative data under these conditions.

There are some recent papers which give an excellent idea of the present possibilities and limitations of probe methods. Sloane and Emeléus (1933) have studied the effect of the length of the probe on the characteristics obtained in the low-voltage arc in A ; Emeléus, Greeves and Montgomery (1936 b) have used probe methods in the negative glow and Faraday dark space in He; Emeléus and Brown (1936 c) and Newman (1938) have pointed out that a particular anomalous form of {current, voltage} probe-characteristic is probably associated with the existence of a positive space-charge for which it may serve as a test. Greeves and Johnston (1936) have studied by means of probes the glow discharge in oxygen in great detail and have discussed the rigour of the methods employed.

(e) *The first Townsend ionization coefficient*

Townsend's early measurements (1915) on the ionization coefficient α/p (where α is the number of ion pairs produced by the passage of one electron through 1 cm. of the gas in the direction of the field) were mainly confined to small electrode distances, low pressures and high values of X/p (>200). The results of much of the work that has been done since are unreliable on account of the newly realized importance of the effect of traces of impurity (*cf.* Penning, 1931 b). Recent experimental work has been done under more carefully controlled conditions, and a number of measurements of α/p over extended ranges in several gases have been made. Sanders (1932, 1933) has made measurements in air for arc lengths of 1 to 7 cm. and values of X/p from 20 to 100. Posin (1936) has obtained α/p for values of X/p from 20 to 1000 in N_2. Kruithof and Penning (1936, 1937 a) have made measurements in pure A for values of X/p from 5 to 1000 and also in Ne–A mixtures. Other work on the subject has also been done (table 6). The very recent work of Bowls (1938) has shown, however, that in many cases, even of the experiments in the last ten years, the values of α/p obtained may be too high as a result of mercury contamination, a trace of which may have the same effect as a trace of A in Ne. Hale (1938, 1939 a has confirmed this effect. In figures 11 *a* and 11 *b* are shown some of the more recent and very probably accurate measurements in a few gases.

In general, $\alpha/p = f(X/p)$ is similar in form for all gases. At low values of X/p (e.g. 20 to 40 in N_2), α/p is proportional to $\exp(A_1 X/p)$, where A_1 is a constant; at high values of X/p (e.g. 200 to 1000 in N_2), α/p varies as $A_3(X/p)^{\frac{1}{2}} - B_3$, where A_3, B_3 are constants. The variation for the intermediate region is given approximately by $(X/p - A_2)^2$, where A_2 is a constant.

There are two reasons why accurate measurements of α/p are of special importance. In the first place, since they are measurements on the fundamental ionization process, the results are of great value in attempts at quantitative explanation of discharge phenomena. For example, these recent accurate data applied to the problem of initiation of the discharge have allowed more confidence to be placed in the underlying theory of the processes taking place (Loeb, 1936). In fact, Posin (1936) from his curves for α/p in N_2, and from measurements on the deviations from the equation

$$i = i_0 \exp(\alpha d), \qquad \ldots\ldots(21)$$

due to secondary processes, was able to compute a curve giving the sparking potential *vs.* pd over a wide range (where p and d are the gas pressure and gap

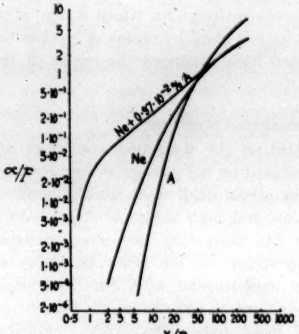

Figure 11 *a*.

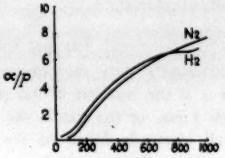

Figure 11 *b*.

distance respectively). The computed values agreed well with experimental results, as can be seen from figure 12. Hale (1939 b) has made similar calculations for H_2. The linking up, in this way, of independent processes to calculate results of a more complicated one is another step in the direction of a complete understanding of discharge processes in terms of the individual impacts.

The second reason why these measurements are important is that the process concerned—although complicated enough—is one of the simplest, as well as one of the most fundamental, in gas discharges. For both these reasons, if we are to have a full explanation of discharge phenomena in terms of the individual processes, the agreement of calculated and measured values of α/p is one of the first essentials. More accurate measurements have now become available through the work mentioned above, and the theoretical developments in connection with the electron-velocity distribution and with electron mobility are immediately applicable to the study of α/p.

The problem of the theoretical derivation of α/p from the individual processes is a difficult one, and, as always, since it has not yet been possible to work out the theory in general terms, the position is made less satisfactory by the lack of data on the individual process parameters. It is possible to express α/p explicitly in terms of X/p, the ionization function (P), the electron velocity distribution function (ρ), the mobility (μ_e) and the ionization potential (V_i), as follows:

$$\frac{\alpha}{p} = \left(\frac{2e}{300m}\right)^{\frac{1}{2}} (X\mu_e)^{-1} \int_{V_i}^{\infty} v^{\frac{1}{2}} \cdot P(v) \cdot \rho(v)\, dv, \qquad \ldots\ldots(22)$$

where e is in e.s.u.

Now, μ_e and $\rho(v)$ are themselves functions of X/p, so that a calculation of α/p from the individual processes requires first of all the calculation of μ_e and $\rho(v)$. Inasmuch as this can be done with fair accuracy, following the work of Smit and

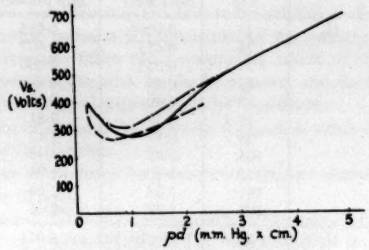

Figure 12. Breakdown voltage for N_2.
Continuous curve—calculated.
Dotted curves—measured.

others, when sufficient data on the collisions are available, α/p could also, with much greater labour, be calculated. The computation is lessened, of course, if measured values of the mobility are used. Since $\rho(v)$ cannot be measured, the only way of saving the labour necessary for calculation here is to assume either Maxwell's or Druyvesteyn's formulae in the regions in which there is some hope of their being valid.

Even using simplifying approximations to the functions, the calculations are very tedious. Druyvesteyn (1932, 1936, 1937), in an attempt to calculate α/p in Ne, assumed a linear variation of the excitation and ionization probabilities and also assumed that the electrons lost no energy at elastic impacts. Values of α/p were calculated for values of X/p from 5 to 30. The agreement with the experimental results seems to be as good as could be expected from the assumptions. Emeléus, Lunt and Meek (1936 a) assumed a Maxwellian velocity distribution to calculate α/p for a number of gases. They used measured values of the mobility and assumed both a linear variation of probability of ionization with

electron velocity and a more accurate form of the variation. The agreement for H_2 (X/p from 10 to 200), air (X/p from 20 to 600), O_2 (X/p from 30 to 120) and N_2 (X/p from 10 to 150) was good within the limits both of experimental and theoretical accuracy. In A, however, there was a whole order of magnitude discrepancy, which may be due to the greater effect of the assumption of the Maxwellian distribution in a monatomic gas on the results. The agreement between measured and calculated values for H_2 and N_2 is shown in table 5.

It is clear, therefore, that we are approaching the stage at which values of α/p computed from atomic data will be available for a number of gases. When this is the case, a great advance in discharge theory will have been made.

Table 5. Measured and calculated values of α/p.
(Emeléus *et al.*, 1936)

	X/p	α/p (calc.)	α/p (meas.)
H_2	13	0·0004	0·015
	21	0·015	0·027
	31·5	0·089	0·090
	43·5	0·23	0·26
	57·5	0·40	0·50
	72·5	0·57	0·81
	88	0·77	1·12
	105	0·95	1·45
	123	1·10	1·75
	161	1·33	2·24
	203	1·47	2·66
N_2	10	$1·54\times10^{-5}$	0·011
	20	$1·09\times10^{-3}$	0·024
	30	$6·19\times10^{-3}$	0·038
	50	0·109	0·082
	86	0·745	0·307
	108	1·25	0·584
	160	2·54	1·51

(f) *The second Townsend coefficient and the initiation of discharges*

During the last ten years the position has been clarified very greatly in connection with the mechanism of the starting of gas discharges. A reasonably clear estimate of the relative importance of the various processes taking place can now be made. Although accurate quantitative data concerning the detailed processes during the initiation of a discharge are, in general, not available, nevertheless sufficient information has now been obtained to allow the important factors to be understood.

It is well known that there are three essential stages in the initiation of a discharge. A few electrons must be produced by some external means, either by radioactivity, cosmic rays or radiation. In practical applications the first two of these are usually adequate. These electrons are accelerated by the applied field causing ionization by collision. The efficiency of this ionization is measured

by the first Townsend ionization coefficient, and is often referred to as the primary process. As a result of this ionization, measurable currents are obtained between the electrodes, but this does not lead to a self-sustaining discharge since the current is still dependent on, and, in fact, usually proportional to, the number of electrons supplied by external means. To produce a self-maintaining discharge, some other processes—usually referred to as the secondary processes—must become operative to supply electrons and make the current independent of those released by external means. The secondary processes come into operation as the applied field is increased. When the number of electrons released near the cathode by the secondary processes is sufficient, the discharge becomes self-sustaining, and the current, unless it is limited by some agency independent of the gas itself, will increase almost without limit, resulting in a spark or some other form of disruptive breakdown.

There has never been much question as to the validity of the assumption that ionization by electron collision is responsible for the primary process. Much discussion has, however, taken place concerning which of the many possible secondary processes is responsible for the breakdown, and the following possible mechanisms of electron production have to be considered:

(*a*) Production of electrons by electrons themselves either in the body of the gas or at the cathode.
(*b*) Production of electrons by positive-ion impact either in the gas itself or at the cathode.
(*c*) Production of electrons by photons either in the gas itself or at the cathode.
(*d*) Action of metastable atoms either in the gas itself or at the cathode.
(*e*) Field emission at the cathode.
(*f*) Space-charge effects.

Some of these may be definitely ruled out; others occur, if at all, only under very exceptional circumstances and are probably not of importance, since in these cases it is likely that other more probable processes would be far more active at the same time. In the first category may be placed (*a*); if, at any stage during the building up of the discharge, the external source supplying the initial electrons were removed, all the electrons already present would be swept to the anode, and, in the absence of any other secondary mechanism, no fresh electrons could be supplied in the cathode region. Moreover, secondary electrons cannot be produced at the cathode surface by electron bombardment since electrons cannot reach the cathode against the opposing field.

In the second category may be placed ionization in the gas by positive ions, and field emission at the cathode. It is shown above that the energies required by positive ions before they become effective ionizing agents are far greater than could be attained under ordinary discharge conditions (see Townsend, 1939; Varney *et al.*, 1940). Moreover, other considerations in connection with the time of development of the discharge lead to the conclusion that ionization by positive ions plays no part in the secondary process; if ionization by positive

ions is the effective secondary process leading to breakdown, times of at least 10^{-5} sec. would be wanted even for short arc lengths. It has been shown, however, that sparks occur in times of the order of 10^{-8} sec. or less (Burawoy, 1926). With regard to field emission, it is unlikely that the very high surface fields required (of the order of 10^7 volts per cm.) could ever be obtained under conditions in which other secondary processes were not far more active (see, however, Flowers (1935)).

Of the other possible processes, it seems now agreed that none of them is in all cases of greater importance than the others. Several may occur simultaneously, and the relative importance of each will depend on the conditions obtaining in the particular case considered.

The action of metastable atoms either in the gas or at the cathode cannot be always an essential part of the secondary process, since the secondary process can occur in cases where metastable atoms are not present. Similarly, at low gas pressures, short gaps and small primary currents, it is unlikely that space-charge distortion of the field can be an important factor. In pure gases at low pressure, photo-ionization in the gas itself is unimportant, since the only mechanism available for the emission of ionizing radiation is the very weak recombination spectrum.

Thus, for the important case of a pure gas at low pressure possessing no metastable states, and with a small value of the primary current, the two most probable secondary processes are :

(1) Release of electrons by impact of photons on the cathode.
(2) Release of electrons by impact of positive ions on the cathode.

It is nearly certain that both these processes are operative under nearly all conditions, even when some of the other processes mentioned above are also active. Under precisely what conditions each is the more important has not yet been settled. Theory does not give much help here, since the equations derived for the breakdown potential on either assumption are similar in form except for the different definition of the constants (Loeb, 1936). The criterion for sparking is given by

$$(1+\gamma)/\gamma = \exp(\alpha d) \qquad \ldots\ldots(23)$$

for electron liberation by positive-ion impact at the cathode, where γ is the number of electrons liberated from the cathode per impact and d is the gap. In the case of photo-ionization at the cathode the criterion is

$$(d+\theta\eta g)/\theta\eta g = \exp[(\alpha-\mu)d], \qquad \ldots\ldots(24)$$

where θ is the number of photons produced by an electron per cm. of discharge path in the direction of the field, μ is the coefficient of absorption of the photons in the gas, η is the fraction of the photons which produce electrons that succeed in leaving the cathode, and g is a geometrical factor representing the fraction of the photons, created in the gas, that reach the cathode. Since μ is small compared with α (except perhaps at high pressures of the order of 1 atmosphere or more),

the experimental accuracy obtainable does not allow differentiation between the two hypotheses. One of the difficulties is that neither the efficiency of low-energy positive ions nor the efficiency of the radiation likely to be emitted during the pre-discharge period, in releasing electrons from a surface, is known. Some recent papers bearing on this question are those of Costa and Raether (1938) and Schade (1938), showing that, at least under some conditions, the photo-effect at the cathode is the more important, and Grigorovici (1939) and Jones (1939 b), who consider that, in their experiments on Hg vapour, the action of positive ions at the cathode is predominant. It is probable that these two actions at the cathode nearly always take place simultaneously during the starting period, and they are usually the most important of the secondary processes.

Distortion of the field by space charge is not usually important at low pressures with short gaps, and with small initial currents. Under all other conditions it probably has some effect. This is indicated experimentally by the low potentials at which sparks in long gaps occur—potentials which give fields (on the assumption of a uniform field) far too low to account for the ionization processes. The results of Varney, White, Loeb and Posin (1935 b), who examined theoretically the effect of space-charge formation on gap breakdown, showed that in certain circumstances the space-charge formation due to primary electron ionization appears to be sufficient to cause breakdown without the assistance of any other secondary process. These conditions involve the pressure, gap distance, applied potential and initial current; and the possibility of the necessary space-charge formation occurring is determined by the form of $f(X/p)$ in the equation $\alpha/p = f(X/p)$. Rogowski (1936) has also examined theoretically the effects of space charge.

Space charge undoubtedly plays an important part in the formation of sparks at large gap distances and at high pressures. Usually it is to be expected that it will be accompanied by release of electrons from the cathode either by impact of positive ions or by photo-electric emission; these processes do not occur, of course, in lightning formation, when space-charge effects are almost certainly predominant.

Photo-ionization in the body of the gas probably becomes effective at high pressures, particularly in the case of impure gases or mixtures of gases. In pure gases, since radiation of wave-length not longer than that corresponding to the ionization potential is required, photo-ionization must be accompanied either by a fairly powerful recombination spectrum or by the spark spectrum. It is likely to be accompanied by the other possible secondary processes. In lightning discharges, and, in fact, in the case of long sparks in air generally, photo-ionization in the air is probably an essential mechanism in the propagation of the streamers (see section (g), on The long spark in air).

That ionization by metastable atoms in the gas itself may be important has been shown by Penning (1931 b), who found a considerable lowering of the starting potential of the low-pressure glow discharges in Ne and He by the addition of a minute trace of A; this process, however, is not of general importance.

In the same way, metastable atoms are capable of releasing electrons from the cathode; their effectiveness in this respect is not very great, since the concentration is relatively small and they are not drawn to the cathode by the field as are the positive ions.

With regard to the effect of cathode material on the breakdown voltage, it seems that, other things being equal, the surface with the lower work function has generally the lower starting voltage. This will depend, of course, upon whether the secondary-emission process at the cathode is made more efficient by a lower work function. It would be expected that this would be so with photo-electric emission, and possibly also in the case of positive-ion bombardment. In long discharges where other processes are more important, the effect will not be so marked.

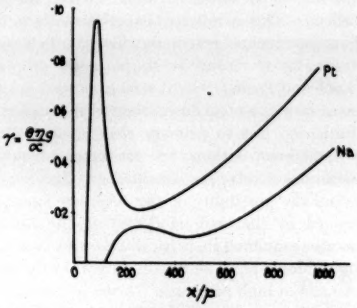

Figure 13.

A number of recent measurements have been made which throw light on this question. Bowls (1938) and Hale (1939 a) have experimented with N_2 and H_2. For the first time, the importance of Hg contamination on the cathode is shown, and it is pointed out that nearly all the earlier studies must have been influenced by this effect. The first and second Townsend ionization coefficients are found as functions of X/p for a Pt cathode and for a Pt cathode contaminated by Hg and by Na. In the latter case, however, the measurements are under suspicion, as the result of the possible formation of a volatile Na compound. The considerable influence of the surface of the cathode on the secondary processes is placed beyond doubt (see also Ehrenkrantz, 1939, and Jones *et al.*, 1938). The curves showing $\gamma\left(=\frac{\theta\eta g}{\alpha}\right)$ *vs.* X/p have a sharp peak for the Pt cathode and a smaller one for the Na-contaminated cathode at X/p about 150 (figure 13). These peaks are not explained, but it is suggested that they are due to some action at the cathode depending on the presence of excited molecules. More recently

Hale (1939 b) has extended his measurements to surfaces of Ni and Al in H_2. α/p is found to be independent of the surface, as, of course, it should be. The values of γ for the two surfaces are not very different, both, however, showing a peak at $X/p = 130$, nearly the same value as in the case of Pt. That γ is not very different for the different metals may be due to the fact that, owing to the presence of hydrogen, the effective surface is the same in each case. Huxford (1939) has also studied the $\{\gamma \textit{ vs. } X/p\}$ relation for Cs–Ag–O surfaces and Ni–Cs surfaces in A. All these measurements are of the first importance in helping towards an understanding of the mechanism, but many more data will be needed before an accurate quantitative theory of the cathode processes to fit all conditions can be worked out.

In practical applications there are two further considerations about which much has still to be discovered. The first of these arises when discharges in tubes long compared with their diameter are used (Seeliger *et al.*, 1938; Mierdel *et al.*, 1937 and 1939). In these cases, all the secondary processes discussed above are relevant, but, in addition, the presence of the insulating glass wall has a very marked effect on the starting voltage. It is known that at least one of the factors concerned is the accumulation of charges on the glass, which distort the field; wide differences in starting voltage can be obtained by varying the capacity to earth of the glass wall and by the presence of conducting layers on the surface. No satisfactory means of controlling these effects have yet been worked out.

The second practical problem which is of importance is the transition of the glow discharge to the arc. When a device is intended to operate at a high current and the circuit conditions are adjusted accordingly, it is sometimes found that, instead of developing an arc discharge, a current of only a few microamperes passes. The current is limited in some way by the surface conditions at the cathode which prevent the development of the arc. Even when the arc starts normally, this type of glow discharge always forms as an intermediate stage after the development of the ordinary secondary processes. The conditions which give rise to rapid and certain transition from this form of glow to the arc are of considerable practical interest. Some of the factors which influence this transition are known; for example, a cathode with a small thermal capacity, which will heat up quickly and provide thermionic emission, is better, other things being equal, than a large one. Moreover, the work function of the surface, if high, tends to delay the transition to the arc stage. The exact conditions which are required, however, are not yet understood, and it may be that the physical state of the surface is important (Maxfield *et al.*, 1938).

An interesting question which Loeb (1936, 1939 a) has discussed recently is whether the usual sparking criteria (equations (23) and (24)) are adequate to represent the breakdown conditions. Since these represent a steady-state condition, they cannot represent the conditions during the breakdown period, and the actual sparking criterion must be more complicated. Thus the actual

sparking potential does depend on the initial supply of electrons (Loeb, 1936). The agreement between observed and calculated {starting voltage *vs. pd*} curves would therefore appear to be to some extent fortuitous, and must not be expected in all cases where the above criteria for sparking are applied.

Recent measurements of interest are those of Klarfield *et al.* (1938 a) and Quinn (1939), who made measurements at pressures less than that corresponding to the Paschen minimum, where the breakdown voltage rises rapidly with decreasing pressure. Slepian and Mason (1937) re-examined experimental data of other authors purporting to verify Paschen's law. By plotting the results in a form designed to utilize the range of d used for constant pd, they found small deviations from Paschen's law in a number of cases. In particular, the results of Cobine and Power (1937) do not suggest that Paschen's law holds for the re-ignition potential of an arc.

(g) *The long spark in air*

The mechanism of spark formation at high pressures and with long gaps is quite different from that at low pressures. At high pressures, space-charge effects and photo-ionization in the body of the gas are of much greater importance than the cathode processes. Under high-pressure conditions it is easier to study visually the actual formation of the discharge than in the case of low pressures, and many excellent observations have been made. These experiments provide valuable data not only in connection with sparking in high pressures, but also on the problem of initiation generally, because there is a more or less gradual transition between conditions existing at high pressure and at low. The work is, of course, stimulated and helped by the similarity to the naturally occurring lightning discharge.

Until recently, the study of sparks was confined to short arc lengths of the order of a few cm. or less. The Kerr electro-optical shutter (Dunnington, 1931) and the cloud chamber (Raether, 1937; Flegler, 1937) have been used to study the formation of the spark in these short gaps. In the last few years, however, much experimental work has been done on the formation of sparks in long gaps (20 cm. or more). This work has been described by Allibone (1938 a) in much greater detail than is possible in this review. The stationary camera and the study of Lichtenberg figures (see, for example, Merrill *et al.*, 1939) provide interesting data on the appearance and formation of the discharges. The most important results, however, have been obtained by the use of applied impulse voltages associated with the cathode-ray oscillograph for the study of the electrical characteristics of breakdown, and with the rotating camera to give a progressive record of the visual spark formation.

The recent work—for example, that of Allibone (1938 a), Allibone and Meek (1938 b, 1938 c), Kip (1938) and Stekolnikov *et al.* (1938)—has made available a full descriptive account of spark formation in inhomogeneous fields. It may be said that, from a descriptive point of view, the division between the old laboratory short sparks and the naturally occurring long spark—lightning—has

been fairly bridged. Theoretically, the position is not quite so happy. More or less satisfactory explanations can be produced for nearly all the phenomena, but a consistent quantitative theory which will embody all the facts has yet to be built up. Excellent attempts in this direction are already being made.

The characteristic feature of spark formation is the propagation of " leader strokes " or streamers which may start from anode or cathode or even in the middle of the gap (White, 1936) and which always precede the main breakdown discharge. It frequently happens that the leader stroke is stepped, that is to say, that several of these streamers may start, progress some distance and then be extinguished one after the other before one succeeds in crossing the gap. When this finally happens, the main breakdown discharge takes place along the low-resistance ionized path. It is in these leader strokes that the principal interest lies, and the theory of their formation is of fundamental importance. About the chief factors in this phenomenon there seems little doubt (see Loeb, 1939 a; Loeb *et al.*, 1939 b). Intense fields which increase enormously the value of α/p are provided by the positive space-charge left by the passage of an electron. This same ionization process produces radiation which supplies other electrons by photo-ionization throughout the body of the gas, and, in particular, in advance of the path taken by the original electron. These electrons in the high fields of the positive space-charge are accelerated, producing further ionization. If the externally applied field and other conditions are such that the ionization produced by the photo-electrons is enough to provide sufficient space-charge to maintain the field at the head of the streamer, it will move forward as an "electron avalanche". This mechanism seems reasonable on general grounds, and with the positive leader, when the electrons must travel towards the point in the opposite direction to the streamer itself, it is difficult to imagine any satisfactory sequence of events other than photo-ionization ahead of the tip which would account for the forward motion. Further evidence in favour of the correctness of this theory is given by consideration of the time lag (i.e. the time between the application of the voltage and the passage of the breakdown current). At high pressures this time is too small to allow a positive ion, or even an electron, to cross the gap completely. Photo-ionization in the body of the gas allows several avalanches to be started more or less simultaneously, each of which has only a small fraction of the gap to cross. A satisfactory explanation for the zig-zag nature of the spark also seems to follow from this picture; a series of avalanches started by photo-ionization, away from the main path, can move towards it and, if the space-charge conditions are favourable, join up with it. The stroke now has the choice of two directions, and it may be easily imagined that frequently the path off the main track offers the lesser resistance.

The whole process is, as the appearance of the sparks would suggest, quite different from that occurring at pressures of only a few mm. Hg. Visual observation at high pressures shows that breakdown occurs along narrow filaments, whereas all the indications, both theoretical and experimental, suggest that at

the lower pressures, where the cathode mechanism is important, practically the whole body of the gas is involved.

At low pressures, although the actual stages of the development of the breakdown are not understood in the sense that it is not yet possible to set up a space-current function dependent on time, there is no doubt that it is a co-operative phenomenon in which many initial electrons are involved. At high pressures, on the other hand, a single electron, under favourable conditions, may suffice to start a train of events each depending on the other and resulting finally in the formation of a spark. Cravath and Loeb (1935 b), Goodlet (1937) and Loeb (1939 a) have discussed the problem theoretically. Very recently, Meek (1940 a) has carried the ideas a step further and developed a quantitative theory to account for the propagation of the streamers. Meek assumed that the streamer develops when the radial field about the positive space-charge attains a value of the order of the external applied field, so that electrons are drawn into the stem of the avalanche and give rise to a conducting filament. Satisfactory agreement is found for the measured and calculated breakdown voltage in air for values of pd of 100 mm. Hg × cm. The time lag calculated (of the order of 10^{-7} sec.) also agrees with that observed. Whether or not the effect of the radial field is as important as Meek's work suggests, there seems no doubt that the general idea of dependence on photo-ionization and space charge is correct. All the observed phenomena can be explained qualitatively on this basis.

In most of the recent experiments the extreme form of inhomogeneous field, namely, point to plane electrode, has been used, as in these conditions the visual observations are more easily made than with homogeneous fields. The process of spark formation differs somewhat according to whether the point is positive or negative, earthed or at the high potential. There are, however, some features common to all cases: the discharge is first seen at the smaller electrode where the field strength is greatest, the value of this field at the point when the discharge first appears being about the same as that necessary to cause breakdown in the case of a homogeneous field. The discharge always proceeds away from the point electrode towards the weaker parts of the field, whether it is positive or negative.

The positive streamers or leader strokes are propagated more readily than the negative ones. Thus, when the point is negative, the voltage at which the spark occurs is nearly twice that when it is positive. The reason for this is that positive space-charge is built up more readily in the region of the anode, and, moreover, the field at the cathode side of a positive space-charge is greater than that at the anode side. For sparks between two points, the breakdown voltage is independent of the polarity and is about half-way between the voltages in the other two cases.

A study of the time lag throws much light on the processes occurring, and in recent years considerable attention has been given to this question. The time lag is composed of two parts: the statistical lag, which depends on the chance of an electron occurring at the proper place to trigger off the primary and secondary

ionization processes, and the formative lag—that is, the time taken for the primary and secondary process to build up into the spark. The statistical time lag may vary from short intervals of the order of 10^{-7} sec. or less up to periods of minutes, depending on the degree of over-voltage and the number of initiatory electrons available. The formative time lag may be as small as 10^{-8} to 10^{-7} sec., but is larger for short gaps (of the order of 1 mm.). Tilles (1934) studied these time lags and found that for low illumination intensities, when the statistical lag was important, they could be represented by a formula of the form

$$\tau = aI^{-c} \exp\left\{-b\left(\frac{\Delta V}{V_s}\right)\right\} \qquad \ldots\ldots(25)$$

where $\Delta V/V_s$ is the fractional over-voltage, I is the primary photo-electric current-density and a, b and c are constants depending on the electrodes, arc length, nature of gas and pressure. The total lag decreases with increasing over-voltage and with increasing initial photo-electric current to a more or less constant value as the statistical lag becomes shorter and the time of the more nearly constant formative lag is reached. White (1936) and Wilson (1936 a) studied the effect of gap length, intensity of illumination and over-voltage in air. They found that for the same time lag, the higher the intensity of illumination the lower the over-voltage. For a given illumination, the time lag decreased as the voltage increased, and time lags as small as 10^{-9} sec. were obtained with over-voltages of the order of 200 per cent. With the short lags obtained (of the order of 10^{-8} sec.) it does not appear possible that either electrons or positive ions could have crossed the gap. This, as already mentioned, provides strong evidence in favour of the view that the simultaneous formation by photo-ionization of space-charge controlled sources of ionization in the body of the gas is responsible for breakdown.

There is not much information available on the variation of the time lag of spark formation with the nature of the gas and pressure (see, however, Schade (1937) and Dunnington (1931)). The formative time lag appears to increase with decreasing pressure, as would be expected, since the conditions approach more nearly to those which obtain when the time of transit of an electron from one electrode to the other is the important factor.

The simplest spark formation in air at atmospheric pressure occurs when the point is positive, the plane is earthed and an appreciable over-voltage with a steep wave front is applied. Then a leader stroke carrying a current of an ampere or so passes from the point to the plane (positive leader stroke), to be followed by the main stroke carrying the full spark current from the negative plane to the point. This travels along nearly the same path as that taken by the leader stroke. If the point is negative and the plane earthed, the leader from the point (negative leader stroke) never reaches the plane, but is met by another leader from the plane. If the plane has no appreciable projection, they meet somewhere near the middle of the gap. If there is a pointed projection on the plane, the leader stroke from the plane is longer.

The leader stroke from the point may be, and usually is, stepped, the composite leader stroke travelling with a velocity of the order of 10^7 cm. per sec.: the individual leaders, following a previously ionized path, may travel at a speed some 100 times this velocity. The velocity of the leader stroke does not depend greatly on whether it originates from the point or the plane.

The time of development of the spark is dependent on the circuit constants and the extent to which the applied voltage exceeds the minimum sparking potential. This is not unexpected, since the form of the voltage wave-front depends on these constants. The composite nature of the leader stroke is also almost certainly dependent on the circuit conditions (Allibone, 1938 a). If the applied voltage has not a steep wave front, it is likely that by the time a streamer, started in the high field near the point, has progressed well away from the electrode, the applied field would not have increased rapidly enough to maintain it. This is probably the explanation of the many leader strokes terminating in space which precede the one that succeeds in bridging the gap. The pre-discharge leaders extinguish one after the other until the applied voltage is not only high enough to start a leader stroke, but also increases rapidly enough to provide the field which will allow it to cross the entire gap.

Allibone (1938 a) has extended his observations with the rotating camera to pressures below atmospheric. Down to about 5 cm. pressure the main features of spark formation in heterogeneous fields seem to persist. As the pressure is reduced, the leader strokes become more sharply defined and more widely separated in time. The explanation is probably associated with the greater difficulty of producing space-charge distortion at the lower pressures.

Further measurements with a number of gases over wide ranges of pressure would provide valuable data which would help towards the formation of a quantitative theory. So much progress has already been made towards understanding spark formation that further quantitative work should soon be forthcoming. Convincing evidence that the study of long-spark formation in the laboratory will ultimately contribute considerably to a complete understanding of lightning formation is given by the fact that photographs of long sparks may be practically indistinguishable in the structure of the flash from actual photographs of lightning (Allibone, 1938 a).

(h) *Negative ions in discharges*

In most commercial discharge devices, negative ions play little part. In the last few years, however, certain types of advertising sign—for example the well known "ripple tubes"—have made use of them. A study of their properties is valuable on general grounds, since a very small concentration in a discharge can have surprisingly large effects. A great deal of the importance of modern technique in the manufacture of discharge devices lies in the care taken to exclude gases and vapours which would lead to the formation of negative ions. The ions are formed in the discharge by electron attachment in a number of gases

such as O_2 or in many organic vapours; they may also be formed by the dissociation of molecules or may even be produced at the electrodes. It seems that their mean lives may be very long—comparable with those of metastable atoms. In a well pumped vacuum device utilizing a discharge through a monatomic gas or metal vapour, if negative ions are present at all it is in such small quantities as to be unimportant.

A very small concentration of negative ions can alter completely all the important properties of a discharge device. The volt-ampere characteristic, the stability of the positive column, and the light output are all affected to a considerable degree. In general, the resistivity of the discharge is increased by the presence of negative ions; the striated positive column and the " ripple " positive column are probably caused by their presence; a small trace of O_2 or H_2O can decrease the light output from a Na or Hg vapour lamp very considerably. The negative ions affect the discharge processes mainly in two ways: the effective collision cross-section of the gas is increased by the process of electron attachment; this accounts for the increased resistance, and may account for the " constricting " effect of negative ions; their low mobility compared with that of electrons allows them to have a major effect in modifying space charge, and the " ripple " effect and striations probably come about in this way.

One of the most useful papers published recently in giving a detailed survey of the present views on the effects caused by negative ions is that of Emeléus and Sayers(1938), in which the results of adding small quantities of Cl to a Ne discharge are studied. The authors describe the effect on the appearance of the glow discharge. They consider the phenomena theoretically, but a comprehensive theoretical treatment of negative ions in discharges has not yet been worked out.

(i) *The positive column*

The positive column is perhaps the simplest as well as the most important region of the discharge. In the absence of disturbing factors, such as the presence of negative ions, the discharge column is usually uniform along and symmetrical about its axis. This is one of the reasons why the uniform positive column has probably been studied more than any other region of the discharge, since its uniformity makes it very suitable for the theoretical and practical study of the fundamental processes. In recent years the positive column has become of great practical importance owing to the growing commercial use of discharge lamps as sources of illumination.

The mechanism of the uniform positive column has been worked out theoretically during the past ten years or so, and is now fairly well understood. It differs considerably according to the gas pressure. There are two main classifications, the low pressure (of the order of 1 mm. Hg) and the high pressure (of the order of 200 mm. Hg or more); quite different theoretical treatments are required in the two cases. There is perhaps a third classification, the very low pressure condition—in which the electron mean-free-path is large compared with, or at

least of the same order as, the dimensions of the tube. This has been considered by Tonks and Langmuir (1929), but is of much less practical importance than the other cases. The simplest and most important low-pressure condition is when the pressure is high enough to make the electron mean-free-path small compared with the tube diameter, but not small enough to cause appreciable lack of radial uniformity in the electron temperature. This condition, for example, holds approximately with the rare gases up to a few mm. Hg pressure. The basis of an adequate theory was laid by Schottky (1924 a, 1924 b), Tonks and Langmuir (1929), Seeliger *et al.* (1931, 1932 a, 1932 b), and Druyvesteyn (1933).

The positive column is characterized by the almost complete absence of space charge (so that the electron and positive-ion concentrations are nearly equal) and by low field strengths. Exactly how and why this state of affairs comes about is not easy to see physically (see, for example, Ollendorf and Seeliger, 1932; Zeleny, 1938). Accepting the fact that it exists, however, the processes occurring in consequence are, in general terms, fairly straightforward. A concentration of electrons exists with a velocity distribution which is often approximately Maxwellian. Departures from this occur at high field-strengths and low current-densities, particularly when an appreciable amount of energy is being lost by radiation or when the Ramsauer effect is important.

Electrons are lost by recombination either in the body of the gas or at the wall. At low pressures and small current-densities nearly all the loss is at the walls, but at higher pressures and larger current-densities the recombination effect in the gas becomes of importance. Positive ions are also lost by diffusion to the walls, and the smaller velocity of the ions is compensated for by a negative charge which is built up at or near the wall of the tube. The formation of this space charge results in a radial electric field which is small near the axis, but which increases towards the walls. This increases the ion drift to the wall and decreases the electron drift, so that the numbers of each lost are equal. The electrons and ions thus removed are replaced by others formed by collision processes in the gas, of which by far the most important is ionization by electron collision.

At low pressures and not too high current-densities the electron temperature is substantially uniform over the cross-section (Druyvesteyn, 1933), but the electron density and current density have a maximum along the axis of the tube, and decrease towards the wall. At a certain stage as the pressure is increased, the electron temperature, or mean electron energy if the distribution is non-Maxwellian, ceases to be uniform over the cross-section, and decreases radially towards the wall, and if the pressure is further increased the column commences to constrict. It is the pressure region well below this to which the low-pressure theory applies. In this low-pressure region the electron temperature is much greater than that of the atoms or positive ions, being usually of the order of several thousand degrees centigrade. The positive-ion energy is only slightly greater than that of the neutral particles, neither being much above the energy corresponding to the ambient temperature (see, for example, Jones, 1935).

For a given gas at a given pressure, if the tube diameter and the total current are specified, all the other parameters of the discharge are thereby fixed. The important parameters are the longitudinal field, the electron density and the electron mean energy. The less important parameters are the current to the walls, the positive-ion density and mean energy. Given an ideal gas, which has a single ionization-potential and no excitation levels, and in which ionization takes place only by electron impact, the characteristics of the positive column can be determined in terms of the current, gas pressure and tube diameter. This is, of course, a very simple picture; in nearly all actual discharges complications are introduced by excitation, collisions of the second kind, absorption of resonance radiation, etc. These complications, however, merely modify and do not essentially change the machinery of the positive column.

Consideration of the diffusion to the walls leads to the equation for the electron concentration,

$$N_r = N_0 J_0(2{\cdot}4\,r/R), \qquad \ldots\ldots(26)$$

where R is the radius of the tube, N_r the concentration at a distance r from the axis, N_0 the concentration on the axis, and J_0 is the Bessel function of the first kind of zero order. It leads also to a relation between the coefficient of ambipolar diffusion (D_a) and the number of electrons produced per electron per sec. (Z):

$$R(Z/D_a)^{\frac{1}{2}} = 2{\cdot}4, \qquad \ldots\ldots(27)$$

where

$$D_a = (D_+\mu_e + D_-\mu_+)/(\mu_+ + \mu_e) \simeq \frac{k}{e}\mu_+ T_e, \qquad \ldots\ldots(28)$$

and D_+ and D_- are the coefficients of diffusion of the positive ions and electrons respectively, μ_+ and μ_e are the mobilities of the positive ions and electrons, k is Boltzmann's constant and T_e is the electron temperature.

Equation (27) is obtained on the assumption that the electron density at the wall is zero. There is not much doubt that this is nearly true (see, for example, Druyvesteyn, 1933, and Yarnold, 1936). It cannot be strictly correct, however, since it leads to an infinite value of the wall potential, but it serves as a first approximation.

If certain assumptions are made, Z and D_a can both be expressed approximately in terms of the electron temperature (T_e), the gas pressure (p) and constants depending on the gas. Equation (27) then gives T_e as a function of p, but independent of the current or voltage gradient:

$$T_e^{\frac{1}{2}} \exp\left(\frac{eV_i}{kT_e}\right) = b_1 p^2, \qquad \ldots\ldots(29)$$

where b_1 is a known gas-dependent constant. Knowing the electron temperature, the potential gradient along the axis can be calculated by balancing the gain of energy by the electrons in the field against the loss at impacts, in the form

$$X = b_2 . f^{\frac{1}{2}} T_e \lambda_e^{-1}, \qquad \ldots\ldots(30)$$

where λ_e is the electron mean-free-path and f is the fraction of the electron energy lost per impact; b_2 is another constant known in terms of the gas parameters.

The electron density is proportional to the current (i). An approximate expression for it is

$$N_0 = i[2\pi e\mu_e X\int_0^R rJ_0(2{\cdot}4r/R)\,dr]^{-1} \qquad \ldots\ldots(31)$$

$$= i[1{\cdot}36e\mu_e XR^2]^{-1}. \qquad \ldots\ldots(32)$$

The theory in this crude form is of more use in giving a clear insight into the mechanism of the positive column than in assisting in making calculations of practical value. It can, however, be refined somewhat at the expense of more lengthy analysis by taking into account exciting collisions, cumulative ionization and using more accurate expressions for Z, μ_e, etc. Where the theory has been checked by experiment, agreement is good enough to show that the right lines are being followed. For example, Mierdel (1938) has calculated the potential

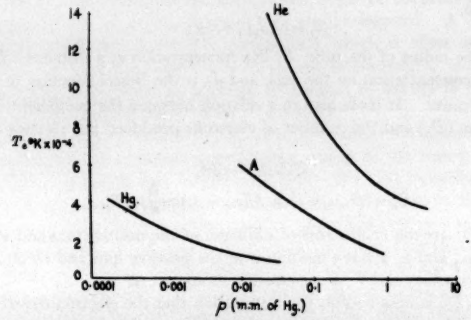

Figure 14. Electron temperature as a function of gas pressure.

gradient in the rare gases using known effective cross-sections to find the electron energy transfer. Klarfeld (1937, 1938 a, 1938 b) has used probe methods to measure accurately the gradient in the rare gases and in Hg vapour as a function of pressure, and correlates the shape of the curves with existing knowledge of the effective cross-section for collision. Reasonably good agreement between theory and experiment is found in the work of both authors.

A few curves from recent papers are given here to show the magnitude of the quantities concerned and the variation with other parameters. According to the above approximate theory, the electron temperature is a function of the gas pressure, but not of the current; figure 14 shows the variation of electron temperature with pressure in a few cases. Figure 15 shows the variation of electron temperature with X/p, and figures 16 a and 16 b show the voltage gradient as a function of pressure.

Funk and Seeliger (1938) have considered theoretically the electron concentration at the walls, which is considered to be zero in the simple theory. The modification made is that the boundary condition at the wall, previously assumed to be $N_R=0$, is given by the diffusion equation

$$\frac{N_R\bar{v}}{4} = -D_a\left(\frac{\partial N_r}{\partial r}\right)_R, \qquad \ldots\ldots(33)$$

where $\bar{v}$ is the mean electron velocity. This leads to finite and reasonable values of the wall potential.

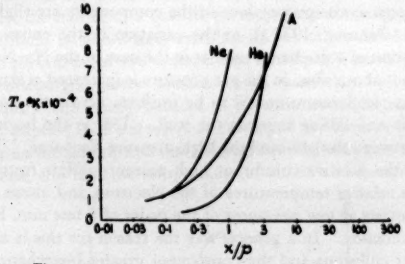

Figure 15. Electron temperature as a function of X/p.

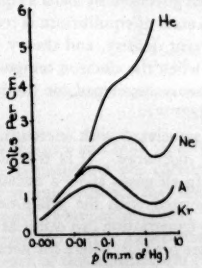

Figure 16 *a*.

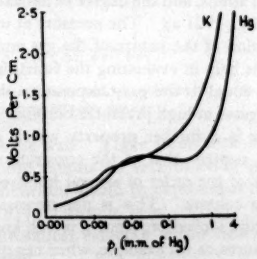

Figure 16 *b*.

Figures 16 *a* and 16 *b*. Voltage gradient as a function of gas pressure.

Klarfeld (1938 c) has examined for Hg vapour the effect of cumulative ionization, which is neglected in the simple theory, and which appears to be important at pressures above about 0·01 mm. Hg, where the above theory applies.

Among other recent work on the low-pressure positive column should be mentioned the following: Hamada (1936) and Ornstein *et al.* (1936), by means of intensity measurements on the band spectrum, have found the gas temperature

in the positive column in N_2 at pressures and current densities which are considered to be below those at which the Maxwellian distribution holds. In the second paper there is an attempt to interpret the results theoretically. Rudolph (1939) has used a Pt wire as a resistance thermometer to measure the temperature rise in the discharge. Alterthum *et al.* (1938) have recently made extensive measurements on the gradient in the positive column in a number of gas mixtures (Ne–A, Ne–Na, Ne–Hg, A–Hg) over a wide range of pressure and current. The study of gas mixtures, in fact, very often proves a most useful tool in the elucidation of the discharge processes. Interesting effects occur, such as the sudden change of colour of a discharge as the proportions of the components are slightly changed (see, for example, Penning, 1934 a), or the variation in the emission spectrum over the cross-section of a discharge tube as in the case of the Na–Ne discharge.

It is pointed out above that, as the gas pressure is increased, a stage is reached when the electron temperature ceases to be uniform across the column, being higher on the axis and falling towards the wall. This is the beginning of the transition stage between the low- and the high-pressure discharge. The essential respect in which the positive column at high pressure differs from that at low pressure is in the relative temperatures of the electrons and atoms. The wide difference which exists at low pressures of the order of a few mm. Hg decreases as the pressure increases. In a general way the reason for this is easily seen in the more frequent collisions and the consequent greater interchange of energy. At a certain pressure, the electron temperature becomes nearly equal to that of the gas atoms, and the degree of ionization is then governed by Saha's equation (Saha, 1920, 1921 a). The pressure at which this state of equilibrium is reached is a function of the nature of the gas and the current density, and theory so far gives little help in evaluating the function. It is when the electron temperature is nearly equal to the gas temperature that the theory developed for discharges through gases at high pressures becomes applicable.

There is a further property which is often associated with increasing gas pressure, namely, that of the constriction of the discharge. It is true that at pressures of the order of several hundred mm. Hg all gases show a constricted discharge column. This is not necessarily associated with the attainment of temperature ionization, since it is well known that constriction can occur at quite low pressures, as, for example, when negative ions are present, where the necessary conditions for temperature ionization do not exist.

There is an intermediate region, of course, between the conditions in which the low-pressure theory holds and those when temperature equilibrium obtains, and this must be treated theoretically as a generalization of the true low-pressure case, without the assumption of uniform temperature over the cross-section (Seeliger, 1932 b and 1932 c).

At pressures above those corresponding to the low-pressure condition, the current tends to concentrate more and more along the axis of the tube, and the increased excitation in this region compared with the outer regions of the column

accounts for the familiar phenomenon of the constricted discharge. Figure 17 shows, for example, the variation in the brightness of the arc across the column of a high-pressure Hg vapour discharge, the gradient being about 40 volts per cm. and the current about 1 amp. The curve for the current-density variation across the column is similar in shape.

Simultaneously with the occurrence of the constriction, the gas temperature increases, the electron temperature decreases, and the state of thermal ionization is approached. Under these conditions, just as in the case of the low-pressure column, given the nature of a gas, its pressure, and the voltage gradient or the current, it should be possible to calculate all the other variables. Most of the work at high pressures (of the order of one atmosphere or more) has been done on the Hg discharge, partly because of its suitability for experimental and theoretical study, and partly because of its commercial importance. Elenbaas in a series of papers (1934 a *et seq.*) has developed the theory of the high-pressure Hg discharge on the assumption that thermal equilibrium exists between electrons, ions and atoms, and that the temperature of atoms and electrons is a maximum

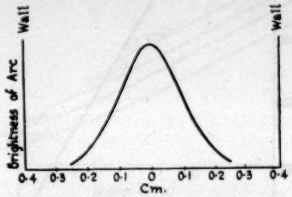

Figure 17. Variation in brightness across luminous column of high-pressure Hg-vapour discharge. (Current 1 amp. ; gradient 40 volts per cm.)

along the axis and falls rapidly towards the wall. There seems little doubt that these assumptions are correct since observations on arcs in other gases give fairly strong evidence that thermal equilibrium would be expected at these high pressures (Mannkopf, 1933; Witte, 1934; Ornstein *et al.*, 1934). There is, however, no direct experimental or theoretical evidence which establishes the existence of this equilibrium beyond doubt.

If these assumptions are correct, it is possible to apply the equation of Saha (1920, 1921 a, 1921 b) to give the degree of ionization in terms of the pressure, temperature and nature of the gas. This relation may be expressed in the form

$$\frac{x^2 p}{1-x^2} = AT^{5/2} \exp\left(-\frac{eV_i}{kT}\right), \qquad \ldots\ldots(34)$$

where A is a known constant, x the degree of ionization, p the pressure, T the gas temperature, and V_i the ionization potential. This formula does not take into account the possibility of the existence of excited states or of radiation equilibrium, and if these considerations are included the formula will contain other factors (see additional references in table 6).

Neglecting excitation this equation gives the degree of ionization in terms of the pressure, temperature and known gas constants. Thus, if the temperature distribution over the cross-section of the column is known, the electron concentration can be found, and from it all the other characteristics of the discharge. The gas-temperature distribution is determined by the condition of energy balance and by the distribution of energy input over the cross section. The energy input $i(r) \cdot X$, where $i(r)$ is the current density at a distance r from the axis of the lamp, must equal the loss of heat by convection, conduction and radiation. Now $i(r) \cdot X$ can be expressed in terms of $T(r)$ and X by means of equation (34), and the mobility equation. The losses by radiation depend on the temperature

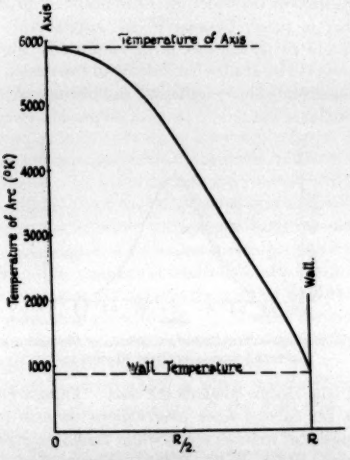

Figure 18. Distribution of temperature over cross-section of high-pressure Hg-vapour discharge.

and pressure, and similarly the conduction and convection losses may be expressed in terms of the temperature and known constants. Theoretically the energy-balance equation could be solved to give the temperature. Actually, however, the equation is too complicated for solution, particularly in respect of the radiation term. Elenbaas (1934 a), therefore, for a particular case measures the temperature at the axis experimentally by an indirect method; a discharge tube contains a quantity of Hg such that for a known wattage input the pressure is just one atmosphere as determined by the balance of pressure inside and outside the bulb when the glass softens; this gives the mean density and pressure so that, assuming a reasonable temperature distribution, the temperature at the axis is found to be

6000° ± 800° K. for a current of 5·5 amp., and a gradient of 8·8 volts per cm. This temperature is of the same order as that found by various methods for arcs in other gases at high pressure (Engel *et al.*, 1931, 1933 a, 1933 b ; Suits, 1935, 1936). The trend of the distribution across the axis is shown in figure 18. Elenbaas (1934 b), using this temperature, calculated for various voltage gradients values of the current, the light-output distribution (and thus the contraction of the discharge) and the total light-output. The good agreement with measured values suggests that the postulated mechanism is correct, and additional evidence for this is obtained from considering the effect produced by the addition of Cd (Elenbaas, 1935 a) and the rare gases (Elenbaas, 1936 b) to the high-pressure mercury discharge. Elenbaas (1935 c) and Heller (1935) have further considered on the same basis the similarity laws for high-pressure discharges.

Barnes and Adams (1938 a, 1938 b) have questioned the validity of Elenbaas's assumption of temperature equilibrium. Their experiments on measurements

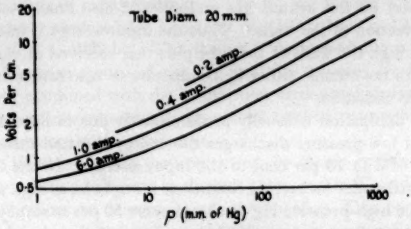

Figure 19. Voltage gradient in the high-pressure Hg-vapour discharge as a function of pressure and current.

of line intensities in the spectrum indicate a considerable difference between the electron temperature and the gas temperature, but no appreciable variation in electron temperature or gas temperature from the axis to the walls. They conclude that the variation in intensity forming the constriction is due to the greater concentration of electrons near the axis, which in turn is due to the radial potential gradient. It does not appear that the question can be finally settled until further information about the probabilities of the various processes taking place in the discharge is available, so that definite theoretical support can be obtained (see, however, Elenbaas, 1939 b).

Other useful work has been done on the high-pressure state during the last few years. Ende (1934) and Elenbaas (1935 f) have made an experimental investigation of the variation of the potential gradient with pressure and current (see figure 19), and have also estimated the losses at the electrodes. Kenty (1938), by photographing the motion of incandescent particles in the discharge, has studied the convection currents, which are of importance both in connection

with an understanding of the energy balance of the positive column and in connection with its stability.

The question of the energy balance of the positive column is one of very great practical importance, since on this depends directly the luminous efficiency of the source and indirectly the volt-ampere characteristic. The electrical power supplied to the tube causes the acceleration of the electrons and positive ions in the field, and is equal to $i \,.\, X$ per unit length of tube, where i is the total current. The energy gained in the gas by the charged particles is lost in several ways: at an inelastic impact, ionization or excitation may take place; if ionization, the electron formed might drift to the wall and recombine there, liberating energy as heat; if excitation, the excited atom might later return to the ground level, liberating one or more quanta of energy, or it might make another impact, passing on some or all of its energy to a neutral atom. The quanta might be transmitted through the walls of the tube or might be absorbed there, causing a further increase in wall temperature. The energy received in various ways (largely by elastic collisions) by the neutral gas molecules is also finally carried by conduction or convection to the walls. Thus the input energy is balanced by that transmitted through the walls of the tube plus that received in various ways by them, and lost to the outside either by conduction or convection or by means of low-temperature radiation.

The energy dissipation is usually predominantly due to impact losses at the walls. In most low-pressure discharges the energy radiated through the walls is of the order of 2 to 10 per cent of the input energy. In the Na discharge, however, operated under favourable conditions it might be as high as 90 per cent. In the case of the high-pressure Hg discharge some 50 per cent may be radiated.

At low pressures Druyvesteyn (1933) has measured the recombination at the walls in a Ne positive column and Druyvesteyn and Warmoltz (1934) have studied experimentally and theoretically the energy balance in the Na discharge with a rare-gas admixture. All the evidence derived from considerations of the energy balance support the conclusion that the mechanism as pictured here is substantially correct. At high pressures the whole Elenbaas theory is based on the energy-balance relations, and more recently Rompe and Schulz (1939 b, 1939 c) have extended this work by considering the effect on the heat-conduction term of the diffusion of excited states and positive ions, and have integrated the energy-balance equation to give a relation between mean values of the parameters.

In two rather important papers Rompe and Schön (1938 a, 1938 d) have studied the dependence of the shape of the volt-ampere characteristic on the magnitude of the amounts of energy dissipated by various processes. The power input is iX and the electron concentration proportional to i/X; therefore, if the power output is known as a function of the electron concentration, the slope of the characteristic is known provided it is assumed that only one dissipative process is operative. Each type of energy loss is examined as a function of the electron concentration, and the shape of the characteristic is thus determined for

the case in which all energy losses are assumed to be of this type. Knowing then, how the various forms of energy loss are influenced by the different variables operating in the discharge, this enables a qualitative estimate to be made of how the shape of the characteristic is likely to vary in different circumstances. The method should be capable of further development.

The series of important papers by Mohler (1932 a *et seq.*) on the Cs discharge has helped considerably to an understanding of the positive column. Most of these papers have been referred to in other places in this review. They have not only yielded fairly complete information on the characteristics of the Cs discharge itself, but have also thrown light on fundamental principles. The experiments were made by various methods over a wide range of conditions ranging from the lower pressure end of the low-pressure range to conditions where it appears that thermal equilibrium exists. Measurements on the potential gradient and other characteristics of the discharge can be explained reasonably well theoretically.

A considerable amount of valuable work has been done on freely burning arcs in various gases, and the positive columns of these have been studied intensively both theoretically and experimentally. The N_2 arc at high pressures has been fairly exhaustively explored by Lamar *et al.* (1936, 1939 a, 1939 b, 1940) and Suits *et al.* (1939 a, 1939 b). The theory is developed along much the same lines as the Elenbaas theory of the Hg positive column, by consideration of energy balance combined with the assumption of thermal ionization. That the mechanism is understood seems to be placed beyond reasonable doubt by the fairly good agreement with experiment.

Penning (1938) calculates the energy lost by elastic collisions, exciting collisions and by ionizing collisions in the rare gases, and in N_2 and H_2 for values of X/p up to 300, when the current is infinitely small. This condition does not apply strictly to the positive column, but the paper is nevertheless of considerable interest in this connection.

§ 4. SPECTRA AND RADIATION PROPERTIES

The emission and absorption spectra of discharges in gases have always been important subjects of research. The reason for this is largely that such study was able to afford much information on the atomic processes, where the conditions were properly controlled; but partly, also, on account of the data which could be obtained on the discharge mechanism itself.

There are many examples of researches that have utilized the study of spectra to elucidate the processes occurring in gas discharges: Brose (1919) made use of the Stark effect to measure field strengths. Fabrikant and Pulver (1934) and Elenbaas (1936 e) studied the concentration of the various excited states in the high-pressure Hg-vapour discharge by measurements of the absorption of the light emitted from one lamp by a second similar one; the population of excited states is fundamental in any consideration of the energy balance of a discharge and of the electron concentration and velocity distribution. Jones (1936 a)

has considered the relation between the electron-velocity distribution and the line intensities in the He positive column. Mohler's work on the recombination spectrum has yielded much valuable information with regard to conditions in the positive column, particularly in the case of Cs. Hamada (1936) has obtained data on electron temperatures in the N_2 positive column by a study of intensity measurements of rotation band spectra.

Although the study of spectra has thus been a useful tool, until the last few years the emission spectrum was definitely considered a by-product of the discharge mechanism proper. In fact, the tendency was to view radiation somewhat reproachfully in so far as it complicated enormously the processes involved and made theoretical interpretation of experimental observations very difficult. Within the last few years, however, the position has changed completely as a result of the increasing use of discharge lamps for advertising signs and for illumination. Here, the luminous efficiency and brightness depend on the line intensities and continuous radiation in the visible part of the spectrum, and the study of how this radiation is related to other fundamental processes in the discharge has become of the greatest importance.

To obtain high luminous efficiency* in a discharge, the electron temperature and concentration should be such that the transitions which result in radiation in the visible part of the spectrum are brought about at the expense of those which do not. This is, of course, only one part of the energy-balance process. It is important also to design the tube so that losses due to other causes such as recombination at the walls are as small as possible. A factor which is of considerable importance in determining the line intensities, and thus the luminous efficiency, is that of self-absorption. This has the effect of prolonging the apparent life of an excited state, and thereby increasing the probability of excitation of the higher levels. Strong self-absorption is desirable, therefore, when the useful radiation results from transitions from higher levels, and undesirable when the useful radiation is mainly resonance, because, in the latter case, excitation to levels other than the resonance levels leads to radiation in a less useful part of the spectrum. The diffusion of resonance radiation in a gas has been treated theoretically by Kenty (1932). The probability of a given quantum of radiation escaping

* If $E(\lambda)d\lambda$ is the energy distribution in the spectrum in watts, and if $S(\lambda)$ is the sensitivity curve for the eye, the lumens radiated are given by

$$L=621\int_0^\infty E(\lambda)S(\lambda)\,d\lambda, \qquad \ldots\ldots(35)$$

and the luminous efficiency in lumens per watt is

$$L/W=621\int_0^\infty E(\lambda)S(\lambda)\,d\lambda\Big/\int_0^\infty E(\lambda)\,d\lambda. \qquad \ldots\ldots(36)$$

Thus, one watt of energy transformed into light of the wave-length at the peak of the visibility curve (5500 A.) would yield 621 lumens. In the Na discharge, if all the energy were transformed into radiation at 5890/96 A., the luminous efficiency would be about 475 L/W. The luminous efficiency of gas-filled tungsten lamps is 15–20 L/W.

from a gas is the important consideration. This probability is a complicated function involving among other things the mean life of the atom in the excited state concerned, and the " mean free-path " of the quantum. It is clear that this will, in turn, depend upon the excitation function of the atom. The problem is greatly complicated by the Doppler broadening of the lines.

Comparatively few gases are suitable as a means of efficient light production. He, CO_2 and N_2 at a few mm. Hg pressure give luminous efficiencies of some 1–5 L/W; Ne and low-pressure Hg vapour give efficiencies rather higher, some 10 L/W. Na and high-pressure Hg give luminous efficiencies as high as 50–80 L/W under practical conditions, far higher efficiencies than can be obtained by any other means.* These two types of discharge have, in recent years, been widely used for illumination, and as a result of the practical importance of these lamps, the properties of the discharge have received much attention in the last ten years. The Na and Hg discharges differ completely in the conditions necessary to obtain high efficiency, and practical design of the lamps therefore presents entirely different problems. In both cases, however, the positive column is the part of the discharge which provides the useful light.

In the Na discharge, nearly all the radiation is in the D lines at 5890/96 A., which are the result of transitions to the ground state from the resonance levels at about 2·1 volts, the ionization potential being 5·12 volts. The problem of obtaining high luminous efficiency therefore is that of preventing loss of energy other than by radiation from the resonance levels. The principal ways in which this energy might be lost are by radiation from higher levels, or by transference of energy from excited atoms either at the wall or in the body of the gas by a collision of the second kind. Excitation to higher levels is increased by increase of current density, which provides more electrons for cumulative excitation, and by increase of pressure, which increases the effective life of atoms excited in the resonance state.

Thus in an efficient Na lamp a low current-density and a low pressure are required. If the pressure is too low, however, the losses to the walls by diffusion increase. There are other factors also involved, such as the velocity distribution of the electrons, which depends in a complicated way on the pressure and current density. Nevertheless, it would be expected, as is in fact the case, that at a given pressure the efficiency would increase with decreasing current-density, and for a given current there would be an optimum pressure. In practical lamps it is always necessary to introduce a rare-gas filling in order to assist starting. This lowers the efficiency somewhat both on account of a certain amount of excitation of the rare gas and also because of energy losses in elastic collisions with the rare gas.

Klarfeld and Taraskov (1936), Krefft, Reger and Rompe (1933) and Druyvesteyn and Warmoltz (1934) have studied experimentally the luminous efficiency

* Low-pressure Hg lamps using the ultra-violet radiation to excite luminescent materials on their walls can give higher luminous efficiencies. A green light can be produced in this way with a luminous efficiency of over 100 L/W.

of the Na discharge. De Groot (1932 a, 1932 b) calculated the light emission in the resonance line and obtained fairly good agreement with measured values. Figures 20 *a* and 20 *b* show Klarfeld and Taraskov's measurements at currents

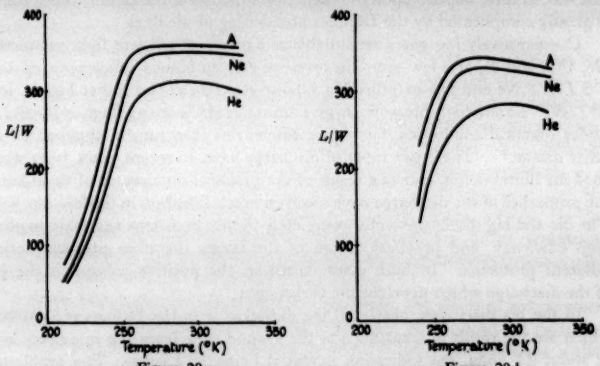

Figure 20 *a*. Figure 20 *b*.

Figures 20 *a* and 20 *b*. Luminous efficiency of the Na-discharge positive column at various temperatures corresponding to the Na-vapour pressure.

Figure 20 *a*. Current 100 ma. ; pressure of rare gas 1 mm. Hg.
Figure 20 *b*. Current 200 ma. ; pressure of rare gas 1 mm. Hg.

of 100 ma. and 200 ma. with A, Ne and He admixture as a function of the temperature corresponding to the Na vapour pressure. Figure 21 from Krefft's paper shows the influence of increasing current with various constant pressures of Na vapour. At constant current, as the Na-vapour pressure increases, the efficiency

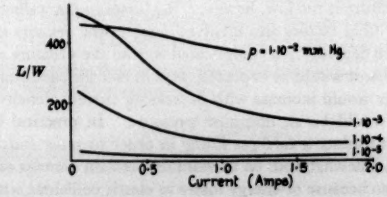

Figure 21. Luminous efficiency of the Na-discharge positive column as a function of the current at various pressures.

first increases as the wall losses become relatively smaller, reaches a maximum at a vapour pressure corresponding to a temperature of about 300° c., then decreases slowly as the re-absorption increases. With increasing current at constant

pressure the results show that the efficiency decreases. The decrease in efficiency due to the presence of rare-gas filling is less with the gas of higher atomic weight. Under experimental conditions, when the required temperature is obtained by means of externally applied heat, efficiencies as high as 300 to 400 L/W are obtained. An essential requirement in practice, however, is that the vapour pressure of the Na has to be produced by heat developed in the lamp by the discharge itself. This means that to increase the pressure it is necessary to increase the current also, and therefore it is not possible to obtain such high efficiencies. In practical lamps the luminous efficiencies obtainable are about 70 L/W.

The low-pressure Hg-vapour discharge has recently been studied by Klarfeld (1937). Different considerations apply here from those in the Na discharge. The resonance radiation, which comprises some 50 per cent of the total input energy, is in the ultra-violet (1850 A. and 2537 A.); the visible lines, of which the strongest are those at 5791 A. and 5461 A., arise from transitions between two excited levels, the 5791 A. line between levels at about 8·8 volts and 6·7 volts ($3\,^1D_2 \rightarrow 2\,^1P_1$) and the 5461 A. line between levels at about 7·7 volts and 5·45 volts ($2\,^3S_1 \rightarrow 2\,^3P_2$). As a result of this, and in contrast with the case of Na-vapour, cumulative excitation and self-absorption are desirable when it is required to utilize the light in the visible part of the spectrum. The stimulation of cumulative excitation is encouraged by the fact that there are metastable levels ($2\,^3P_2$, $2\,^3P_0$) at 5·45 volts and 4·7 volts. At constant current, as the pressure increases, the efficiency increases also, until a pressure of the order of 0·1 mm. Hg (the exact value depending on current and tube diameter) is reached. Above this, the efficiency decreases as the effectiveness of the metastable atoms is reduced by frequent collisions with other atoms, and the energy losses at the walls increase. At a pressure of 2 to 3 mm. Hg a minimum is reached above which the efficiency again increases as the high-pressure region is approached. Here, the discharge column begins to contract from the wall, and the degree of ionization becomes dependent almost entirely on the gas temperature as the electron and gas temperatures approach equilibrium. The population of the higher levels now rises with increasing pressure and current density, and the intensities of the lines in the visible increase at the expense of the others. In contrast with the Na discharge in which the highest luminous efficiency is obtained at a pressure of some 10^{-2} mm. Hg, the Hg-vapour discharge attains its highest efficiency (some 70 L/W) at pressures of the order of several atmospheres. Figure 22 shows the efficiency of the Hg-vapour discharge as a function of pressure.

The high-pressure Hg spectrum has been studied intensively in recent years, partly on account of its practical importance and partly on account of the interest attaching to spectra under conditions of the very high pressures obtainable in Hg vapour. As the pressure increases from a fraction of a millimetre to about one atmosphere, the relative intensities of the lines change (the yellow lines increasing in intensity compared with the blue), and a certain amount of continuous radiation appears. The result of this modification of the spectrum is

to change the colour of the light from blue, characteristic of the low-pressure condition, to one which appears nearly white, although—as the colour-rendering properties show—the radiation is not by any means white. The theory of the continuous spectrum is not yet completely understood (Unsöld, 1938 ; Rössler. 1939; Elenbaas, 1939 a). It is associated both with the electron recombination spectrum and with molecular spectra originating from the loosely bound molecules which may be expected to occur at high pressures.

At still higher pressures (greater than one atmosphere) the lines are considerably broadened, due to fairly well understood effects (see White, 1934), and the study of the degree of broadening of the various lines as a function of current-density and pressure is proving of considerable interest (Rompe *et al.*, 1938 c; Schulz, 1938). In addition, the continuous spectrum becomes very prominent until at pressures of the order of 100 atmospheres the broadened lines merge into the continuous background and an approach to ordinary temperature radiation

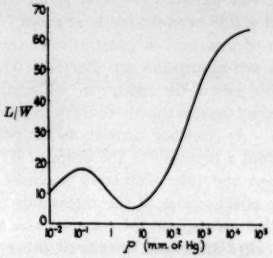

Figure 22. The efficiency of the Hg-vapour discharge as a function of pressure. (Current 4 amp. ; tube diameter 27 mm.)

is obtained (Elenbaas, 1936 d). Figures 23 *a* and 23 *b*, from Elenbaas's paper, show how the intensity of the continuous spectrum increases with increasing pressure and current density. A study of the discharge column in this condition is of great fundamental interest. It represents a condition intermediate between that of a gas radiating the normal line spectrum and a black-body radiator, and may be considered as an approximation to an incandescent body of low thermal emissivity at a temperature of some 5000° C. to 6000° C. The low thermal emissivity is a consequence of the low density compared with that of solid bodies.

A further recent development in connection with the high-pressure Hg discharge is the addition of Tl metal to obtain still higher efficiencies (British Patent 509,801, 1938). Tl has a resonance line at 5350 A., and it is found that as the vapour pressure of Tl in the high-pressure Hg column is increased, the luminous efficiency increases also, both the Hg and Tl lines showing strongly. The efficiency appears to increase up to a vapour pressure of Tl corresponding to 800° to 1000° C., after which there is some indication of a decrease. Figure 24

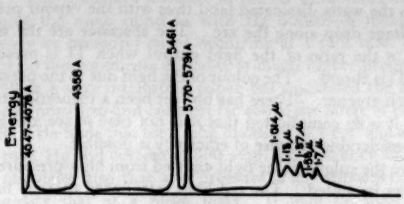

Figure 23 *a*.

Tube diameter 4·3 mm. Current 0·4 amp.
Voltage gradient 120 volts per cm. Pressure 20 atmospheres.

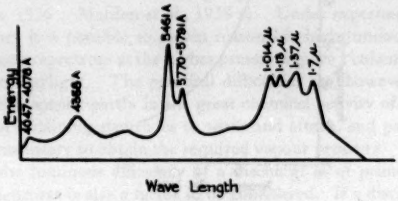

Figure 23 *b*.

Tube diameter 1 mm. Current 1·1 amp.
Voltage gradient 800 volts per cm. Pressure 200 atmospheres.

Figures 23 *a* and 23 *b*.—Distribution of energy in the spectrum of the high-pressure Hg-vapour discharge.

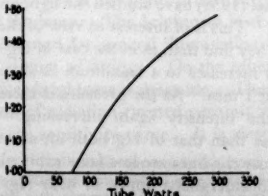

Figure 24. The effect of the addition of Tl vapour to the Hg-vapour discharge. η is the ratio of the luminous efficiency with Tl present to the efficiency with Tl absent. (Tube diameter 8 mm. Voltage gradient 50 volts per cm.)

shows for a particular tube how the gain in efficiency due to the presence of the Tl varies with the watts dissipated (and thus with the vapour pressure of Tl) for a constant voltage drop along the arc. The abscissae are the watts dissipated; the ordinates η the ratio of the light output when Tl is present to the light output when Tl is absent. The colour of the light due to the presence of the Tl is, of course, much greener. There has not yet been a theoretical treatment of this problem, and it is so complicated that it is not easy to see on general grounds why such a considerable increase of efficiency is possible.

To improve the colour of the light emitted from high-pressure Hg discharges, Cd and Zn may be added. The red lines introduced thereby have the effect of improving the colour, but the luminous efficiency is reduced. Figure 25 shows how the increase in percentage of red light in the spectrum * obtained by introducing different amounts of Cd and Zn into the discharge affects the efficiency. Elenbaas (1935 a) has investigated the effect of the addition of Cd theoretically

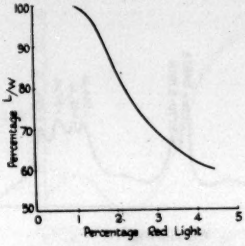

Figure 25. Effect on the luminous efficiency of adding Cd and Zn to the high-pressure Hg-vapour discharge.

and experimentally. He finds that the voltage gradient is increased although the ionization potential of Cd is lower than that of Hg. This is explainable in view of the fact that Cd lowers the temperature of the discharge and thus the degree of ionization.

Fabrikant and Kanel (1935) have studied the light emission from a Cd-vapour discharge without Hg. This is of interest in view of the similar structures of the Cd and Hg atoms. They find that, as in the case of Hg vapour, as the pressure is increased the intensity increases to a maximum at about 0·1 mm. Hg, then falls to a minimum at about 1 mm. As the pressure is increased further the column begins to contract, the intensity again increasing. The total light output obtainable is much less than that of Hg both on account of the lower vapour pressure and also because the lines are less favourably placed in the spectrum.

In Ne the same considerations apply as with Hg. The visible lines are radiated between excitation levels—a group at about 18·5 volts and a group at

* Percentage transmitted by Wratten Filter No. 25. For comparison, the percentage red in daylight is approximately 15 on the same scale.

about 16·6 volts. The resonance lines are in the far ultra-violet at about 740 A. At lower pressures the same trend as with Hg occurs—a maximum in the {luminous efficiency *vs.* pressure} curve occurring at 1 to 2 mm. Hg. pressure. There seem to be no measurements available at much higher pressures of Ne.

Mohler (1939) in a very interesting paper has calculated the characteristics of a Cs discharge at pressures and currents far above the range of present experimental technique. He finds that for a tube diameter of 1 cm. and at a Cs-vapour pressure of 175 mm. Hg and current of 300 amp. the efficiency should be some 180 L/W. He calculates also that at a pressure of one atmosphere the spectral distribution is nearly that of a black body. It may be some time before these conditions are realized experimentally, but if and when they are it would be another example of the stage intermediate between that of a gas emitting a line spectrum and an incandescent body, and the important commercial applications are obvious.

The use of Te vapour in the semi-high-pressure condition has been considered as a possible means of obtaining high luminous efficiency with a good colour (Rompe, 1936 ; Marden *et al.*, 1938 a). Under experimental conditions in the laboratory it is possible to obtain reasonably high luminous efficiencies, and the continuous spectrum at the higher pressures gives a colour which approximates to that of daylight. The practical difficulties are, however, very serious. These difficulties consist partly in the great chemical activity of Te vapour and the difficulty of obtaining electrodes to withstand attack, and partly in the very high loadings necessary to obtain the required vapour pressure.

Although the luminous efficiency of a discharge is of primary importance, the surface brightness is also a factor to be considered. If a discharge column is symmetrical about its axis, the surface brightness is related to the luminous efficiency by the equation

$$L = R \cdot 4\pi \iint_S B\,dS, \qquad \ldots\ldots(37)$$

where L is the light output in lumens, $\iint_S B\,dS$ is the brightness measured in stilb * integrated over the surface viewed at any point at right angles to the axis, and R is the reduction factor † of the lamp. The brightness is of considerable practical importance in lamps to be used for general illumination because a low surface brightness facilitates the design of fittings. On the other hand, for projection purposes a very high surface brightness is desirable. The brightness of a source is measured by the amount of radiation emitted per unit area of the surface, due regard being paid to the eye sensitivity-curve. It is clear that to obtain a high

* 1 stilb = 1 candle per cm^2. For comparison it may be noted that the brightness of the filament of the ordinary tungsten-filament lamp used for general lighting service is about 500 stilb. The maximum brightness of the high current-density arc—the brightest practical source so far available—is about 80,000 stilb.

† $R = \dfrac{\text{Average spherical candle power}}{\text{Directional candle power}}$ (Walsh, 1926.)

brightness it is necessary to dissipate a large wattage in a small space. It is clear also that the conditions essential for high brightness are not necessarily those required to obtain a high luminous efficiency. For example, the brightness of the positive column may usually be increased by decreasing the diameter of the tube, but this frequently results in a lower luminous efficiency. The brightness of the positive columns of the usual form of low-pressure discharge lamp such as the Na or Ne lamps are of the order of only a few stilb. It has recently been found, however, that the high-pressure Hg column has remarkable possibilities in this direction, and brightnesses of the order of 100,000 stilb have been obtained experimentally.

The brightness is not usually uniform over the surface of a discharge column (see, for example, figure 17). It is known that the luminous efficiency of the Hg-vapour discharge continues to increase, or at least does not appreciably decrease, up to pressures of several hundred atmospheres and current densities

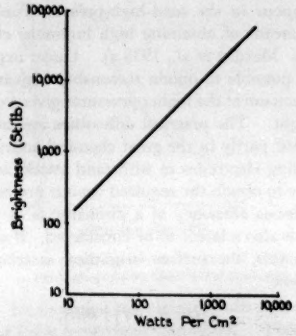

Figure 26. Relation between the wattage dissipated per cm² of surface of a high-pressure Hg-vapour arc and the surface brightness.

of hundreds of amps. per cm². Thus, since the column can be continually constricted by increasing the pressure, and the current density can be increased to very high values, an extremely large dissipation of energy at high luminous efficiency can be obtained from a small surface. Figure 26 shows for a wide range the relation between the energy dissipated per cm² of surface and the surface brightness for the Hg-vapour discharge column. It can be seen that there is no evidence of the curve saturating at the highest values at present attained. The fact that the brightness seems to be independent of the arc voltage and current and to depend only on the wattage dissipated is additional evidence in favour of the thermal ionization mechanism.

It is interesting to speculate as to the theoretical limit of brightness obtainable with the Hg-vapour source. If line radiation is considered, most of the levels

which supply the visible radiation are at about 8 volts, corresponding to a temperature of some 80,000° K. The temperature of the column at the highest values of the brightness shown in figure 26 is probably less than 10,000° K. Thus it would seem that far higher temperatures of the column would be required before the line radiation reached its maximum value. But at these high temperatures and pressures, the source becomes less like a gas radiating a line spectrum and more like a black body. Figure 27 shows the approximate value of the brightness of a black body for temperatures up to 80,000° K. Thus, whichever

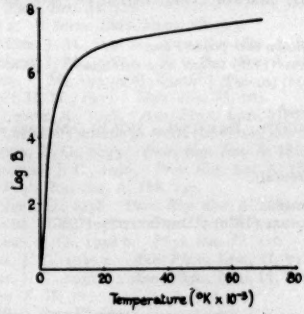

Figure 27. Brightness of a black body (*B*) in stilb as a function of the temperature.

mechanism is the more important in these circumstances, it would seem clear that the theoretical limit of brightness is extremely high provided the conditions of a gas discharge still hold. What actually would happen at these extremely high pressures and under these temperature conditions is difficult to see, and must be left for some fascinating experimental work in the future.

Table 6. Additional references

Critical potentials for electron collisions

Smith (1931) (Hg) ; Tate and Smith (1932 b) (molecular gases) ; Tate and Smith (1934) (alkali metal vapours and rare gases) ; Whiddington and Woodroofe (1935) (rare gases) ; Löhner (1935 a) (rare gases).

Scattering and total cross-section for electron collisions

Brode (1929) (alkali metal vapours); Kollath (1930 b) (N_2) ; Hughes and McMillen (1932) (A) ; Tate and Palmer (1932 a) (Hg) ; Childs and Massey (1933 a, 1933 b) (Hg, Cd, Zn).

Excitation functions for electron collisions

Hanle and Schnaffernicht (1930) (Hg) ; Larché (1931) (Cd, Zn) ; Thieme (1932) (He, Hg, N_2) ; Seiler (1933) (Hg) ; Maier-Leibnitz (1935) (rare gases).

Cross-section for ionization by electron collisions

Compton and Van Voorhis (1926) (rare gases and molecular gases) ; Smith (1930) (rare gases) ; Bleakney (1930 b) (H_2) ; Bleakney (1930 c) (rare gases) ; Smith (1931) (Hg) ; Tate and Palmer (1932 a) (Hg).

Theory of electron collisions
ALLIS and MORSE (1931) ; FEENBERG (1932 a, 1932 b) ; MORSE and ALLIS (1933).

Ionization and ionization functions for collisions by positive ions
SUTTON and MOUZON (1930, 1931) (K^+–He, alkali ions–rare gases) ; BEECK and MOUZON (1931 a, 1931 b) (alkali ions–rare gases) ; WOLF (1932) (A^+–A) ; NORDMEYER (1933) (K^+–A) ; FRISCHE (1933) (K^+–rare gases and molecular gases) ; MOUZON and SMITH (1935) (Mg^+–Ne, A) ; DAVIES (1936) (K^+–He).

Excitation by collision with positive ions
JONES (1933) (L^+, K^+–Hg) ; HOLZER (1934) (alkali ions–rare gases) ; HENLE *et al.* (1936) (alkali ions–Hg); MAURER (1936, 1937 a) (K^+–He, Li^+–He) ; MOSER (1937) (Hg^+–Hg).

Total cross-section for collision with positive ions
RAMSAUER and BEECK (1928) (alkali ions–rare gases) ; KENNARD (1928) (alkali ions–H_2, He, A) ; BEECK (1934) (survey) ; WOLF (1936).

Collisions of the second kind
HANLE and LARCHÉ (1933) ; PIKE (1936 a, 1936 b) ; MAURER and MEHNERT (1937 b).

Recombination
HARPER (1940) (theoretical).

Electron velocity distribution
THOMAS (1928) ; LANDAU (1936) ; DRUYVESTEYN (1938).

Electron mobility.
ELENBAAS (1938).

Probes
HEATLEY (1937) ; SPIWAK and REICHRUDEL (1938) ; FRAGSTEIN and ARNDT (1938).

The first Townsend ionization coefficient
GLOTOV (1937) ; MORALEW (1937) ; GOSSERIES (1939).

Initiation of discharges
KLEMPERER (1930) ; FRICKE (1933) ; PENNING and ADDINK (1934 b) ; PENNING (1934 c) ; KÖHLER (1936) ; FLEGLER *et al.* (1936) ; JURRIAANSE *et al.* (1936) ; BÜTTNER (1939) ; JONES *et al.* (1939 a) ; LOEB and MEEK (1940) ; MEEK (1940 b).

Negative ions
MASSEY and SMITH (1936) ; ARNOT and MILLIGAN (1936, 1937) ;

Positive column (low pressure)
KILLIAN (1930) (Hg) ; LÜBCKE (1933) (Hg) ; GROOS (1934) (A) ; DORGILO *et al.* (1935) (Ne–A, Ne–Hg) ; SIEBERTZ (1936) (rare gas–Hg mixtures) ; LUNT and MEEK (1936) (energy balance) ; v. ENGEL (1936) (Hg) ; STEENBECK (1936) (H_2, D_2) ; WHITE (1937) (Hg) ; SOMMERMEYER (1937) (energy balance).

Positive column (high pressure)
MILNE (1921) (thermal ionization) ; RUSSELL (1922) (thermal ionization) ; FOWLER (1923) (thermal ionization) ; MILNE (1924) (thermal ionization) ; SAHA (1935) (thermal ionization) ; ELENBAAS (1936 c) (turbulent convection) ; ELENBAAS (1937 b) (Zn and Cd) ; SRIVASTAVA (1938) (thermal ionization).

Spectra and radiation properties
DE GROOT and DRUYVESTEYN (1932 a) (diffusion of resonance radiation); DE GROOT (1933) (diffusion of resonance radiation) ; DE GROOT (1934) (diffusion of resonance radiation) ;

ELENBAAS and DE GROOT (1935 g) (high-pressure Hg) ; KREFFT, LARCHÉ and RÖSSLER (1936) (high-pressure Hg) ; KREFFT, RÖSSLER and RÖTTENAUER (1937) (high-pressure Hg) ; FABRIKANT (1938) (excited states) ; RIEWE and ROMPE (1938) (excited states) ; RÖSSLER (1938) (high-pressure Hg) ; KERN (1938 a) (high-pressure Hg) ; MARDEN, BEESE and MEISTER (1938) (high-pressure Hg) ; ROMPE and SCHULZ (1938 *b*) (high-pressure Hg) ; KOCH (1938) (high-pressure Hg) ; KERN and SCHULZ (1938 b) (high-pressure Hg) ; RÖSSLER and SCHÖNHERR (1938 c) (Hg) ; ROMPE (1939 a) (high-pressure Hg).

REFERENCES

ALLEN, H. W., 1937. *Phys. Rev.* **52**, 707.
ALLIBONE, T. E., 1938 a. *J. Instn. Elect. Engrs*, **82**, 513.
ALLIBONE, T. E. and MEEK, J. M., 1938 b. *Proc. Roy. Soc.* A, **166**, 97.
ALLIBONE, T. E. and MEEK, J. M., 1938 c. *Proc. Roy. Soc.* A, **169**, 246.
ALLIS, W. P. and MORSE, P. M., 1931. *Z. Phys.* **70**, 567.
ALLIS, W. P. and ALLEN, H. W., 1937. *Phys. Rev.* **52**, 703.
ALTHERTHUM, H. and LOMPE, A., 1938. *Ann. Phys., Lpz.*, **31**, 1.
ARNOT, F. L., 1933. *Collision Processes in Gases* (Methuen and Co., Ltd.).
ARNOT, F. L., and BAINES, G. O., 1935. *Proc. Roy. Soc.* A, **151**, 256.
ARNOT, F. L. and MILLIGAN, J. C., 1936. *Proc. Roy. Soc.* A, **156**, 538.
ARNOT, F. L., 1937. *Proc. Roy. Soc.* A, **158**, 137.
ARNOT, F. L. and BECKETT, C., 1938. *Proc. Roy. Soc.* A, **168**, 103.
BARNES, B. T. and ADAMS, E. Q., 1938 a. *Phys. Rev.* **53**, 545.
BARNES, B. T. and ADAMS, E. Q., 1938 b. *Phys. Rev.* **53**, 576.
BEECK, O. and MAUZON, J. C., 1931 a. *Ann. Phys., Lpz.*, **11**, 737.
BEECK, O. and MAUZON, J. C., 1931 b. *Ann. Phys., Lpz.*, **11**, 858.
BEECK, O., 1934. *Phys. Z.* **35**, 36.
BELL, M. E., 1939. *Phys. Rev.* **55**, 201.
BLEAKNEY, W., 1930 a. *Phys. Rev.* **35**, 139.
BLEAKNEY, W., 1930 b. *Phys. Rev.* **35**, 1181.
BLEAKNEY, W., 1930 c. *Phys. Rev.* **36**, 1303.
BORGNIS, F., 1938. *Ann. Phys., Lpz.*, **31**, 497.
BOWLS, W. E., 1938. *Phys. Rev.* **53**, 293.
BRACEFIELD, C. J., 1932. *Phys. Rev.* **42**, 11.
BRACEFIELD, C. J., 1933. *Phys. Rev.* **43**, 785.
BRACEWELL, K. H., 1938. *Phys. Rev.* **54**, 639.
BRAMHALL, E. H., 1931. *Proc. Camb. Phil. Soc.* **27**, 421.
BRAMHALL, E. H., 1932. *Phil. Mag.* **13**, 682.
BRATA, L., 1933. *Proc. Roy. Soc.* A, **141**, 454.
BRODE, R. B., 1929. *Phys. Rev.* **34**, 673.
BRODE, R. B., 1933. *Rev. Mod. Phys.* **5**, 257.
BROSE, E., 1919. *Ann. Phys., Lpz.*, **58**, 731.
BROXON, J. W. and MERIDETH, G. T., 1938 a. *Phys. Rev.* **54**, 1.
BROXON, J. W. and MERIDETH, G. T., 1938 b. *Phys. Rev.* **54**, 605.
BURAWOY, O., 1926. *Arch. Elektrotech.* **16**, 186.
BÜTTNER, H., 1939. *Z. Phys.* **111**, 750.
CHILDS, E. C. and MASSEY, H. S. W., 1933 a. *Proc. Roy. Soc.* A, **141**, 473.
CHILDS, E. C. and MASSEY, H. S. W., 1933 b. *Proc. Roy. Soc.* A, **142**, 509.
CHRISTOPH, W., 1935. *Ann. Phys., Lpz.*, **23**, 51.
COBINE, J. D. and POWER, R. B., 1937. *J. Appl. Phys.* **8**, 287.
COMPTON, K. T., 1923 a. *Phys. Rev.* **22**, 333.
COMPTON, K. T., 1923 b. *Phys. Rev.* **22**, 432.
COMPTON, K. T. and VAN VOORHIS, C. C., 1926. *Phys. Rev.* **27**, 724.
COMPTON, K. T. and LANGMUIR, I., 1930. *Rev. Mod. Phys.* **2**, 221.

COSTA, H. and RAETHER, H., 1938. *Naturwissenschaften*, **26**, 593.
CRAVATH, A. M., 1935 a. *Phys. Rev.* **47**, 254.
CRAVATH, A. M. and LOEB, L. B., 1935 b. *Physics*, **6**, 125.
DAVIES, A. C., 1936. *Proc. Roy. Soc.* A, **155**, 123.
DAVYDOV, B., 1935. *Phys. Z. Sowjet*, **8**, 59.
DAVYDOV, B. and ZMANOVSKAJA, L., 1936 a. *Tech. Phys. U.S.S.R.* **3**, 715.
DAVYDOV, B., 1936 b. *Phys. Z. Sowjet*, **9**, 433.
DAVYDOV, B., 1937. *Phys. Z. Sowjet*, **12**, 269.
DORGILO, H. B., ALTING, H. and BOERS, C. J., 1935. *Physica*, **2**, 959.
DRUYVESTEYN, M. J., 1930 a. *Z. Phys.* 64, 781.
DRUYVESTEYN, M. J., 1930 b. *Physica*, **10**, 61.
DRUYVESTEYN, M. J., 1932. *Phys. Z.* **33**, 822.
DRUYVESTEYN, M. J., 1933. *Z. Phys.* **81**, 571.
DRUYVESTEYN, M. J. and WARMOLTZ, N., 1934. *Phil. Mag.* **17**, 1.
DRUYVESTEYN, M. J., 1936. *Physica*, **3**, 65.
DRUYVESTEYN, M. J., 1937. *Physica*, **4**, 464.
DRUYVESTEYN, M. J., 1938. *Physica*, **5**, 561.
DUFFENDACK, O. S., and THOMSON, K., 1933. *Phys. Rev.* **43**, 106.
DUFFENDACK, O. S. and OWENS, J. S., 1934. *Phys. Rev.* **46**, 417.
DUFFENDACK, O. S., and GRAN W. H., 1937. *Phys. Rev.* **51**, 804.
DUNNINGTON, F. G., 1931. *Phys. Rev.* **38**, 1535.
EHRENKRANTZ, F., 1939. *Phys. Rev.* **55**, 219.
ELENBAAS, W., 1934 a. *Physica*, **1**, 211.
ELENBAAS, W., 1934 b. *Physica*, **1**, 673.
ELENBAAS, W., 1935 a. *Physica*, **2**, 45.
ELENBAAS, W., 1935 b. *Physica*, **2**, 155.
ELENBAAS, W., 1935 c. *Physica*, **2**, 169.
ELENBAAS, W., 1935 d. *Physica*, **2**, 757.
ELENBAAS, W., 1935 e. *Physica*, **2**, 763.
ELENBAAS, W., 1935 f. *Physica*, **2**, 787.
ELENBAAS W. and DE GROOT, W., 1935 g. *Physica*, **2**, 807.
ELENBAAS, W., 1936 a. *Physica*, **3**, 12.
ELENBAAS, W., 1936 b. *Physica*, **3**, 219.
ELENBAAS, W., 1936 c. *Physica*, **3**, 484.
ELENBAAS, W., 1936 d. *Physica*, **3**, 859.
ELENBAAS, W., 1936 e. *Physica*, **3**, 947.
ELENBAAS, W., 1937 a. *Physica*, **4**, 279.
ELENBAAS, W., 1937 b. *Physica*, **4**, 413.
ELENBAAS, W., 1937 c. *Physica*, **4**, 747.
ELENBAAS, W., 1938. *Physica*, **5**, 568.
ELENBAAS, W., 1939 a. *Physica*, **6**, 299.
ELENBAAS, W., 1939 b. *Phys. Rev.* **55**, 294.
EMELÉUS, K. G., LUNT, R. W. and MEEK, C. A., 1936 a. *Proc. Roy. Soc.* A, **156**, 394.
EMELÉUS, K. G., GREEVES, F. D. and MONTGOMERY, E., 1936 b. *Proc. Roy. Irish Acad.* **43**, 35.
EMELÉUS, K. G. and BROWN, W. L., 1936 c. *Phil. Mag.* **22**, 898.
EMELÉUS, K. G. and SAYERS, J., 1938. *Proc. Roy. Irish Acad.* **44**, 87.
ENDE, W., 1934. *Z. tech. Phys.* **15**, 601.
VON ENGEL, A. and STEENBECK, M., 1931. *Wiss. Veröff. Siemens-Konz.* **10**, 155.
VON ENGEL, A. and STEENBECK, M., 1933 a. *Wiss. Veröff. Siemens-Konz.* **12**, 74.
VON ENGEL, A. and STEENBECK, M., 1933 b. *Wiss. Veröff. Siemens-Konz.* **12**, 89.
VON ENGEL, A., 1936. *Z. tech. Phys.* **17**, 404.
FABRIKANT, W. A. and PULVER, W. L., 1934. *Phys. Z. Sowjet*, **6**, 521.
FABRIKANT, W. A. and KANEL, A. S., 1935. *Comptes Rendus (Doklady) U.S.S.R.* **3**, 257.

FABRIKANT, V., 1938. *Tech. Phys. U.S.S.R.* **5**, 864.
FEENBERG, E., 1932 a. *Phys. Rev.* **40**, 40.
FEENBERG, E., 1932 b. *Phys. Rev.* **42**, 17.
FERCHMIN, A. and FRISCH, S., 1936. *Phys. Z. Sowjet*, **9**, 466.
FLEGLER, E. and RAETHER, H., 1936. *Z. Phys.* **103**, 315.
FLEGLER, E., 1937. *Elektrotech. Z.* **58**, 1262.
FLOWERS, T. W., 1935. *Phys. Rev.* **48**, 954.
FOWLER, R. H., 1923. *Phil. Mag.* **45**, 1.
VON FRAGSTEIN, C. and ARNDT, M., 1938. *Ann. Phys., Lpz.*, **33**, 532.
FRICKE, H., 1933. *Z. Phys.* **86**, 464.
FRISCHE, C. A., 1933. *Phys. Rev.* **43**, 160.
FUNK, W. and SKELIGER, R., 1938. *Z. Phys.* **110**, 271.
FÜRHMANN, H., 1939. *Ann. Phys., Lpz.*, **34**, 625.
GABOR, D., 1932. *Z. tech. Phys.* **13**, 560.
GABOR, D., 1933 a. *Phys. Z.* **34**, 38.
GABOR, D., 1933 b. *Z. Phys.* **84**, 474.
GARDNER, M. E., 1938. *Phys. Rev.* **53**, 75.
GLOTOV, I. I., 1937. *Phys. Z. Sowjet.* **12**, 256.
GOODLET, B. L., 1937. *J. Instn. Elect. Engrs*, **81**, 1.
GORSCHKOW, M. and MASLAKOWERY, J., 1936. *Tech. Phys. U.S.S.R.* **3**, 824.
GOSSERIES, A., 1939. *Physica*, **6**, 458.
GREEVES, F. D., and McF. JOHNSTON, J. E., 1936. *Phil. Mag.* **21**, 659.
GRIGOROVICI, R., 1939. *Z. Phys.* **111**, 596.
GROOS, O., 1934. *Z. Phys.* **88**, 741.
DE GROOT, W. and DRUYVESTEYN, M. J., 1932 a. *Physica*, **12**, 153.
DE GROOT, W., 1932 b. *Physica*, **12**, 289.
DE GROOT, W., 1933. *Physica*, **13**, 41.
DE GROOT, W., 1934. *Physica*, **1**, 28.
GÜNTHERSCHULZE, A. and BÄR, W., 1938 a. *Z. Phys.* **109**, 121.
GÜNTHERSCHULZE, A., BÄR, W. and WINTER, A., 1938 b. *Z. Phys.* **111**, 208.
GVOSDOVER, S. D., 1937. *Phys. Z. Sowjet*, **12**, 164.
HAFT, G., 1933. *Z. Phys.* **82**, 73.
HALE, D. H., 1938. *Phys. Rev.* **54**, 241.
HALE, D. H., 1939 a. *Phys. Rev.* **55**, 815.
HALE, D. H., 1939 b. *Phys. Rev.* **56**, 1199.
HAMADA, T., 1936. *Proc. K. Akad. wet. Amst.* **39**, 50.
HANLE, W. and SCHAFFERNICHT, W., 1930. *Ann. Phys., Lpz.*, **6**, 905.
HANLE, W. and LARCHÉ, K., 1933. *Z. Phys.* **85**, 548.
HANLE, W., 1934. *Die Physik*, **2**, 85.
HANLE, W., 1938. *Die Physik*, **6**, 33.
HARPER, W. R., 1940. *Phil. Mag.* **29**, 434.
HASELTINE, W. R., 1939. *J. Math. Phys.* **18**, 174.
HASSÉ, H. R. and COOK, W. R., 1929. *Proc. Roy. Soc.* A, **125**, 196.
HASSÉ, H. R. and COOK, W. R., 1931. *Phil. Mag.* **12**, 554.
HEATLEY, A. H., 1937. *Phys. Rev.* **52**, 235.
HELLER, G., 1935. *Physics*, **6**, 389.
HENLE, O. and MAURER, W., 1936. *Phys. Z.* **37**, 659.
HERSHEY, A. V., 1939 a. *Phys. Rev.* **56**, 908.
HERSHEY, A. V., 1939 b. *Phys. Rev.* **56**, 916.
HERTZ, G., 1925. *Z. Phys.* **32**, 298.
HOLZER, R. E., 1934. *Phys. Rev.* **46**, 280.
HUGHES, A. L. and MCMILLEN, J. H., 1932. *Phys. Rev.* **39**, 585.
HUXFORD, W. S., 1939. *Phys. Rev.* **55**, 754.
HUXLEY, L. G. H., 1937. *Phil. Mag.* **23**, 208.

JONES, E. J., 1933. *Phys. Rev.* **44**, 707.
JONES, F. L., 1935. *Proc. Phys. Soc.* **47**, 74.
JONES, F. L., 1936 a. *Proc. Phys. Soc.* **48**, 513.
JONES, F. L. and WILLOTT, W. H., 1936 b. *Proc. Phys. Soc.* **48**, 830.
JONES, F. L. and GALLOWAY, W. R., 1938. *Proc. Phys. Soc.* **50**, 207.
JONES, F. L. and HENDERSON, J. P., 1939 a. *Phil. Mag.* **28**, 185.
JONES, F. L., 1939 b. *Phil. Mag.* **28**, 192.
JURRIAANSE, T. and DRUYVESTYN, M. J., 1936. *Physica*, **3**, 825.
KENNARD, R. B., 1928. *Phys. Rev.* **31**, 423.
KENTY, C., 1928. *Phys. Rev.* **32**, 624.
KENTY, C., 1932. *Phys. Rev.* **42**, 823.
KENTY, C., 1933. *Phys. Rev.* **43**, 181.
KENTY, C., 1938. *J. Appl. Phys.* **9**, 53.
KERN, J., 1938 a. *Z. tech. Phys.* **19**, 249.
KERN, J. and SCHULZ, P., 1938 b. *Z. Phys.* **111**, 454.
KILLIAN, T. J.,1930. *Phys. Rev.* **35**, 1238.
KIP, A. F., 1938. *Phys. Rev.* **54**, 139.
KLARFELD, B. and TARASKOV, I., 1936. *Tech. Phys. U.S.S.R.* **3**, 881.
KLARFELD, B., 1937. *Tech. Phys. U.S.S.R.* **4**, 44.
KLARFELD, B. and GUSJEVA, L., 1938 a. *Tech. Phys. U.S.S.R.* **5**, 425.
KLARFELD, B., 1938 b. *Tech. Phys. U.S.S.R.* **5**, 725.
KLARFELD, B., 1938 c. *Tech. Phys. U.S.S.R.* **5**, 911.
KLEMPERER, O., 1930. *Z. tech. Phys.* **19**, 270.
KOCH, B., 1938. *Ann. Phys., Lpz.*, **33**, 335 (1938).
KÖHLER, A., 1936. *Arch. Elektrotech.* **30**, 528.
KOLLATH, R., 1930 a. *Phys. Z.* **31**, 985.
KOLLATH, R., 1930 b. *Ann. Phys., Lpz.*, **5**, 1017.
KOLLATH, R., 1932. *Ann. Phys., Lpz.*, **15**, 485.
KRAUS, P., 1937. *Ann. Phys., Lpz.*, **29**, 449.
KREFFT, H., REGER, M. and ROMPE, R., 1933. *Z. tech. Phys.* **14**, 242.
KREFFT, H., LARCHÉ, K. and RÖSSLER, F., 1936. *Z. tech. Phys.* **17**, 374.
KREFFT, H., RÖSSLER, F. and RÜTTENAUER, A., 1937. *Z. tech. Phys.* **18**, 20.
KRUITHOF, A. A. and PENNING, F. M., 1936. *Physica*, **3**, 515.
KRUITHOF, A. A. and PENNING, F. M., 1937 a. *Physica*, **4**, 430.
KRUITHOF, A. A. and DRUYVESTEYN, M. J., 1937 b. *Physica*, **4**, 450.
LAMAR, E. S., 1936. *Phys. Rev.* **49**, 861.
LAMAR, E. S., STONE, A. M. and COMPTON, K. T., 1939 a. *Phys. Rev.* **55**, 1145.
LAMAR, E. S., STONE, A. M. and COMPTON, K. T., 1939 b. *Phys. Rev.* **55**, 1235.
LAMAR, E. S. and STONE, A. M., 1940. *Phys. Rev.* **57**, 212.
LANDAU, L., 1936. *Phys. Z. Sowjet*, **10**, 154.
LANGEVIN, P., 1903 a. *Ann. Chim.* (*Phys.*) **28**, 289.
LANGEVIN, P., 1903 b. *Ann. Chim.* (*Phys.*) **28**, 433.
LANGEVIN, P., 1905. *Ann. Chim.* (*Phys.*) **5**, 245.
LANGMUIR, I. and MOTT-SMITH, H., 1923 a. *Gen. Elect. Rev.* **26**, 731.
LANGMUIR, I. and MOTT-SMITH, H., 1923 b. *Gen. Elect. Rev.* **27**, 449, 583, 616, 726, 810.
LANGMUIR, I., 1925. *Phys. Rev.* **26**, 583.
LARCHÉ, K., 1931. *Z. Phys.* **67**, 440.
LEES, J. H., 1932. *Proc. Roy. Soc.* A, **137**, 173.
LIPSON, H. C. and MITCHELL, A. C. G., 1935. *Phys. Rev.* **48**, 625.
LOEB, L. B., 1936. *Rev. Mod. Phys.* **8**, 267.
LOEB, L. B., 1939 a. *Fundamental Processes of Electrical Discharges in Gases* (New York: John Wiley and Sons).
LOEB, L. B. and KIP, A. F., 1939 b. *J. Appl. Phys.* **10**, 142.
LOEB, L. B. and MEEK, J. M., 1940. *J. Appl. Phys.* **11**, 438.

Löhner, H., 1935 a. *Ann. Phys., Lpz.*, **22**, 81.
Löhner, H., 1935 b. *Ann. Phys., Lpz.*, **24**, 349.
Lübcke, E., 1933. *Z. tech. Phys.* **14**, 297.
Lunt, R. W. and Meek, C. A., 1936. *Proc. Roy. Soc.* A, **157**, 146.
Mächler, W., 1936 a. *Z. Phys.* **104**, 1.
Mächler, W., 1936 b. *Phys. Z.* **37**, 211.
Maier-Leibnitz, H. and Sponer, H., 1935. *Z. Phys.* **95**, 499.
Manley, J. H. and Duffendack, O. S., 1935 *Phys. Rev.* **47**, 56.
Mannkopff, R., 1933. *Z. Phys.* **86**, 161.
Marden, J. W., Beese, N. C. and Meister, G., 1938 a. *J. Franklin Inst.* **225**, 45.
Marden, J. W., Beese, N. C. and Meister, G., 1938 b. *Trans. Illum. Engng Soc., N.Y.*, **33**, 147.
Mason, R. C., 1937. *Phys. Rev.* **51**, 28.
Massey, H. S. W. and Mohr, C. B. O., 1931 a. *Proc. Roy. Soc.* A, **132**, 605.
Massey, H. S. W. and Mohr, C. B. O., 1931 b. *Proc. Roy. Soc.* A, **136**, 289.
Massey, H. S. W. and Mohr, C. B. O., 1933 a. *Proc. Roy. Soc.* A, **139**, 187.
Massey, H. S. W. and Mohr, C. B. O., 1933 b. *Proc. Roy. Soc.* A, **140**, 613.
Massey, H. S. W. and Smith, R. A., 1933 c. *Proc. Roy. Soc.* A, **142**, 142.
Massey, H. S. W. and Mohr, C. B. O., 1934. *Proc. Roy. Soc.* A, **146**, 880.
Massey, H. S. W. and Smith, R. A., 1936. *Proc. Roy. Soc.* A, **155**, 472.
Massey, H. S. W., 1938. *Negative Ions* (Cambridge University Press).
Maurer, W., 1936. *Z. Phys.* **101**, 323.
Maurer, W., 1937 a. *Z. Phys.* **104**, 658.
Maurer, W. and Mehnert, K., 1937 b. *Z. Phys.* **106**, 453.
Maurer, W., 1939. *Phys. Z.* **40**, 161.
Maxfield, F. A. and Fredenhall, G. L., 1938. *J. Appl. Phys.* **9**, 600.
Meek, J. M., 1940 a. *Phys. Rev.* **57**, 722.
Meek, J. M., 1940 b. *Proc. Phys. Soc.* **52**, 547.
Merrill, F. H. and von Hippel, A., 1939. *J. Appl. Phys.* **10**, 873.
Mierdel, G. and Steenbeck, M., 1937. *Z. Phys.* **106**, 311.
Mierdel, G., 1938. *Wiss. Veröff. Siemens-Konz.* **17**, 277.
Mierdel, G., 1939. *Wiss. Veröff. Siemens-Konz.* **18**, 68.
Milne, E. A., 1921. *Observatory*, **44**, 269.
Milne, E. A., 1924. *Phil. Mag.* **47**, 209.
Mitchell, J. H. and Ridler, K. E. W., 1934. *Proc. Roy. Soc.* A, **146**, 911.
Möhler, F. L. and Boeckner, C., 1929. *Bur. Stand. J. Res., Wash.*, **2**, 489.
Mohler, F. L., 1932 a. *Bur. Stand. J. Res., Wash.*, **9**, 25.
Möhler, F. L., 1932 b. *Bur. Stand. J. Res., Wash.*, **9**, 493.
Möhler, F. L., 1933 a. *Bur. Stand. J. Res., Wash.*, **10**, 357.
Möhler, F. L., 1933 b. *Bur. Stand. J. Res., Wash.*, **10**, 771.
Möhler, F. L., 1936 a. *Bur. Stand. J. Res., Wash.*, **16**, 227.
Möhler, F. L., 1936 b. *Bur. Stand. J. Res., Wash.*, **17**, 45.
Möhler, F. L., 1936 c. *Bur. Stand. J. Res., Wash.*, **17**, 849.
Möhler, F. L., 1937 a. *Bur. Stand. J. Res., Wash.*, **19**, 447.
Möhler, F. L., 1937 b. *Bur. Stand. J. Res., Wash.*, **19**, 559.
Möhler, F. L., 1938 a. *Bur. Stand. J. Res., Wash.*, **21**, 697.
Möhler, F. L., 1938 b. *Bur. Stand. J. Res., Wash.*, **21**, 873.
Möhler, F. L., 1939. *J. Opt. Soc. Amer.* **29**, 152.
Moralew, S., 1937. *Phys. Z. Sowjet*, **12**, 89.
Morse, M. and Allis, W. P., 1933. *Phys. Rev.* **44**, 269.
Morse, M., Allis, W. P. and Lamar, E. S., 1935. *Phys. Rev.* **48**, 412.
Moser, H., 1937. *Ann. Phys., Lpz.*, **28**, 97.
Mouzon, J. C. and Smith, N. H., 1935. *Phys. Rev.* **48**, 420.
Newman, F. H., 1938. *Phil. Mag.* **26**, 719.

NIELSEN, R. A., and BRADBURY N. E., 1936 a. *Phys. Rev.* **49**, 388.
NIELSEN, R. A., 1936 b. *Phys. Rev.* **50**, 950.
NIELSEN, R. A. and BRADBURY, N. E., 1937. *Phys. Rev.* **51**, 69.
NORDMEYER, M., 1933. *Ann. Phys., Lpz.*, **16**, 706.
NOTTINGHAM, W. B., 1928. *J. Franklin Inst.* **206**, 43.
NOTTINGHAM, W. B., 1929. *J. Franklin Inst.* **207**, 299.
NOTTINGHAM, W. B., 1939. *Phys. Rev.* **55**, 203.
OLLENDORF, F. and SEELIGER, R., 1932. *Phys. Z.* **33**, 577.
ONSAGER, L., 1938. *Phys. Rev.* **54**, 554.
ORNSTEIN, L. S. and BRINKMAN, H., 1934. *Physica*, **1**, 797.
ORNSTEIN, L. S., BRINKMAN, H. and HAMADA, T., 1936. *Proc. K. Akad. wet. Amst.* **39**, 315.
PAETON, H., 1939. *Z. Phys.* **111**, 770.
PEARCE, A. F., 1936. *Proc. Roy. Soc.* A, **155**, 490.
PENNEY, W. G., 1932. *Phys. Rev.* **39**, 467.
PENNING, F. M., 1928. *Physica*, **8**, 13.
PENNING, F. M., 1931 a. *Z. Phys.* **72**, 338.
PENNING, F. M., 1931 b. *Phil. Mag.* **11**, 961.
PENNING, F. M., 1932. *Z. Phys.* **78**, 454.
PENNING, F. M., 1934 a. *Physica*, **1**, 763.
PENNING, F. M. and ADDINK, C. C. T., 1934 b. *Physica*, **1**, 1028.
PENNING, F. M., 1934 c. *Physica*, **1**, 1028.
PENNING, F. M., 1938. *Physica*, **5**, 286.
PIKE, E. W., 1936 a. *Phys. Rev.* **49**, 513.
PIKE, E. W., 1936 b. *Phys. Rev.* **49**, 515.
POSIN, D. Q., 1936. *Phys. Rev.* **50**, 650.
POWELL, C. F. and BRATA, L., 1932. *Proc. Roy. Soc.* A, **138**, 117.
QUINN, R. B., 1939. *Phys. Rev.* **55**, 482.
RAETHER, H., 1937. *Z. Phys.* **107**, 91.
RAETHER, H., 1938. *Z. Phys.* **110**, 611.
RAMSAUER, C. and BEECK, O., 1928. *Ann. Phys., Lpz.*, **87**, 1.
RIEWE, K. H. and ROMPE, R., 1938. *Z. Phys.* **111**, 79.
ROGOWSKI, W., 1936. *Z. Phys.* **100**, 1.
ROMPE, R., 1936. *Z. Phys.* **101**, 214.
ROMPE, R. and SCHÖN, M., 1938 a. *Z. Phys.* **108**, 265.
ROMPE, R. and SCHULZ, P., 1938 b. *Z. Phys.* **108**, 654.
ROMPE, R. and SCHULZ, P., 1938 c. *Z. Phys.* **110**, 223.
ROMPE, R. and SCHÖN, M., 1938 d. *Z. Phys.* **111**, 345.
ROMPE, R. and SCHÖNHEER, F., 1938 e. *Z. tech. Phys.* **19**, 588.
ROMPE, R., 1939 a. *Z. Phys.* **112**, 369.
ROMPE, R. and SCHULZ, P., 1939 b. *Z. Phys.* **112**, 691.
ROMPE, R. and SCHULZ, P., 1939 c. *Z. Phys.* **113**, 10.
RÖSSLER, F., 1938. *Z. Phys.* **110**, 352.
RÖSSLER, F., 1939. *Z. Phys.* **112**, 667.
RUDOLPH, G., 1939. *Z. Phys.* **111**, 523.
RUSSELL, H. N., 1922. *Astrophys. J.* **55**, 119.
SAHA, M. N., 1920. *Phil. Mag.* **40**, 472.
SAHA, M. N., 1921 a. *Z. Phys.* **6**, 40.
SAHA, M. N., 1921 b. *Proc. Roy. Soc.* A, **99**, 135.
SAHA, M. N. and SRIVASTAVA, B. N., 1935. *A Treatise on Heat*, chapter XVIII (The India Press, Ltd.).
SANDERS, F. H., 1932. *Phys. Rev.* **41**, 667.
SANDERS, F. H., 1933. *Phys. Rev.* **44**, 1020.
SAVARD, J., 1933. *J. Phys. Radium*, **4**, 650.

SAVARD, J., 1934 a. *J. Phys. Radium*, **5**, 27.
SAVARD, J., 1934 b. *J. Phys. Radium*, **5**, 283.
SAVARD, J., 1934 c. *Physica*, **1**, 321.
SAVARD, J., DE HEMPTINNE, M. and CAPRON, P., 1937 a. *C.R. Acad. Sci., Paris*, **204**, 354.
SAVARD, J. DE HEMPTINNE, M. and CAPRON, P., 1937 b. *C.R. Acad. Sci., Paris*, **204**, 1039.
SAVARD, J. and DE HEMPTINNE, M., 1939. *J. Phys. Radium*, **10**, 30.
SAYERS, J., 1938. *Proc. Roy. Soc.* A, **169**, 83.
SCHADE, R., 1937. *Z. Phys.* **104**, 487.
SCHADE, R., 1938. *Z. Phys.* **111**, 437.
SCHMIDT, O., 1934 a. *Ann. Phys., Lpz.*, **21**, 241.
SCHMIDT, O., 1934 b. *Ann. Phys., Lpz.*, **21**, 268.
SCHMIDT, O., 1936. *Ann. Phys., Lpz.*, **25**, 92.
SCHOTTKY, W., 1924 a. *Phys. Z.* **25**, 342.
SCHOTTKY, W., 1924 b. *Phys. Z.* **25**, 635.
SCHOTTKY, W., 1925. *Z. Phys.* **31**, 163.
SCHULZ, P., 1938. *Z. tech. Phys.* **19**, 585.
SEELIGER, R. and HIRCHERT, R., 1931. *Ann. Phys., Lpz.*, **11**, 817.
SEELIGER, R., 1932 a. *Phys. Z.* **33**, 273.
SEELIGER, R., 1932 b. *Phys. Z.* **33**, 313.
SEELIGER, R., 1932 c. *Phys. Z.* **33**, 577.
SEELIGER, R. and BOCK, K., 1938. *Z. Phys.* **110**, 717.
SEILER, R., 1933. *Z. Phys.* **83**, 789.
SIEBERTZ, K., 1936. *Ann. Phys., Lpz.*, **27**, 421.
SLEPIAN, J. and MASON, R. C., 1937. *J. Appl. Phys.* **8**, 619.
SLOANE, R. H. and EMELÉUS, K. G., 1933. *Phys. Rev.* **44**, 333.
SLOANE, R. H. and MCGREGOR, E. I. R., 1934. *Phil. Mag.* **18**, 193.
SMIT, J. A., 1937. *Physica*, **3**, 543.
SMITH, P. T., 1930. *Phys. Rev.* **36**, 1293.
SMITH, P. T., 1931. *Phys. Rev.* **37**, 808.
SMITH, R. A., 1938. *Proc. Roy. Soc.* A, **168**, 19.
SNAVELY, B. L., 1937. *Phys. Rev.* **52**, 174.
SOMMERMEYER, K., 1937. *Ann. Phys., Lpz.*, **28**, 240.
SPIWAK, G. and REICHRUDEL, E., 1936. *Physica*, **3**, 301.
SPIWAK, G. and REICHRUDEL, E., 1938. *Tech. Phys. U.S.S.R.* **5**, 715.
SRIVASTAVA, B. N., 1938. *Proc. Roy. Soc.* A, **167**, 484.
STEENBECK, M., 1936. *Z. tech. Phys.* **17**, 397.
STEKOLNIKOV, J. and BELJAKOV, A., 1938. *Tech. Phys. U.S.S.R.* **5**, 251.
SUITS, C. G., 1935. *Physics*, **6**, 190, 196, 315.
SUITS, C. G., 1936. *Gen. Elect. Rev.* **39**, 194.
SUITS, C. G., 1939 a. *Phys. Rev.* **55**, 561.
SUITS, C. G. and PORITSKY, H., 1939 b. *Phys. Rev.* **55**, 1184.
SUITS, C. G., 1939 c. *J. Appl. Phys.* **10**, 730.
SUTTON, R. M. and MOUZON, J. C., 1930. *Phys. Rev.* **35**, 694.
SUTTON, R. M. and MOUZON, J. C., 1931. *Phys. Rev.* **37**, 379.
TATE, J. T. and PALMER, R. R., 1932 a. *Phys. Rev.* **40**, 731.
TATE, J. T. and SMITH, P. T., 1932 b. *Phys. Rev.* **39**, 270.
TATE, J. T. and SMITH, P. T., 1934. *Phys. Rev.* **46**, 773.
THOMAS, L. H., 1928. *Proc. Roy. Soc.* A, **121**, 464.
THOMSON, J. J., 1924. *Phil. Mag.* **47**, 337.
THIEME, O., 1932. *Z. Phys.* **78**, 412.
TILLES, A., 1934. *Phys. Rev.* **46**, 1015.
TONKS, L. and LANGMUIR, I., 1929. *Phys. Rev.* **34**, 876.
TONKS, L., 1938. *Phys. Rev.* **54**, 634.
TOWNSEND, J. S., 1915. *Electricity in Gases* (Oxford University Press).

TOWNSEND, J. S. and YARNOLD, G. D., 1934. *Phil. Mag.* **18**, 594.
TOWNSEND, J. S., 1939. *Phil. Mag.* **28**, 111.
TYNDALL, A. M. and POWELL, F. C., 1932. *Proc. Roy. Soc.* A, **136**, 145.
TYNDALL, A. M. and PEARCE, A. F., 1935. *Proc. Roy. Soc.* A, **149**, 426.
TYNDALL, A. M., 1938. *The Mobility of Positive Ions in Gases* (Cambridge University Press).
UNSÖLD, A., 1938. *Ann. Phys., Lpz.*, **33**, 607.
UYTERHOEVEN, W., 1938. *Elektrische Gasentladungslampen* (Berlin : Julius Springer).
VAN GORCUM, A. H., 1936. *Physica*, **3**, 207.
VARNEY, R. N., 1935 a. *Phys. Rev.* **47**, 483.
VARNEY, R. N., WHITE, H. J., LOEB, L. B. and POSIN, D. Q., 1935 b. *Phys. Rev.* **48**, 818.
VARNEY, R. N. and LOEB, L. B., 1935 c. *Phys. Rev.* **48**, 822.
VARNEY, R. N., 1936 a. *Phys. Rev.* **50**, 159.
VARNEY, R. N. and COLE, A. C., 1936 b. *Phys. Rev.* **50**, 261.
VARNEY, R. N., COLE, A. C., and GARDNER M. E., 1937. *Phys. Rev.* **52**, 526.
VARNEY, R. N., LOEB, L. B. and HASELTINE, W. R., 1940. *Phil. Mag.* **29**, 379.
WALSH, J. W. T., 1926. *Photometry*, chapter IV (London : Constable and Co., Ltd.).
WAYLAND, H., 1937. *Phys. Rev.* **52**, 31.
WHIDDINGTON, R. and WOODROOFE, E. G., 1935. *Phil. Mag.* **20**, 1109.
WHITE, H. E., 1934. *Introduction to Atomic Spectra*, chapter XXI (McGraw-Hill).
WHITE, H. J., 1936. *Phys. Rev.* **49**, 507.
WHITE, J., 1937. *Tech. Phys. U.S.S.R.* **4**, 138.
WILSON, R. R., 1936 a. *Phys. Rev.* **49**, 201.
WILSON, R. R., 1936 b. *Phys Rev.* **50**, 1082.
WITTE, H., 1934. *Z. Phys.* **88**, 415.
WOLF, F., 1932. *Z. Phys.* **74**, 575.
WOLF, F., 1936. *Ann. Phys., Lpz.*, **27**, 543.
WOLF, F., 1937. *Ann. Phys., Lpz.*, **28**, 438.
WOLF, F., 1938. *Ann. Phys., Lpz.*, **31**, 561.
YARNOLD, G. D., 1936. *Phil. Mag.* **22**, 988.
ZANSTRA, H., 1935. *Physica*, **2**, 817.
ZELENY, J., 1938. *J. Franklin Inst.* **226**, 35.

SOME INTERACTIONS OF GASES WITH METALS AND CRYSTALLINE SOLIDS

By J. K. ROBERTS

General introduction

When a gas atom or molecule approaches a solid surface it may either rebound after suffering an elastic or inelastic collision or it may remain for an appreciable time adsorbed on the surface. We shall consider here both types of interaction. In Part 1 we shall discuss phenomena connected with the interchange of energy in collisions between gases and solids and in Part 2 the structure of the adsorbed films found when the gas molecules adhere to the solid. Actually we shall find that the two groups of phenomena are not unconnected, in that energy interchange between gas and solid is profoundly affected by the presence or absence of adsorbed films, so that the measurement of this interchange under defined conditions provides a method for the quantitative study of the properties of adsorbed films which otherwise would be inaccessible to observation.

PART 1. THE INTERCHANGE OF ENERGY BETWEEN GASES AND SOLIDS

§1. THEORY OF HEAT LOSS FROM A FINE WIRE IN A GAS AT LOW PRESSURE

We shall discuss the interchanges of energy between gases and solids in terms of the accommodation coefficient of Knudsen, which is defined as follows:

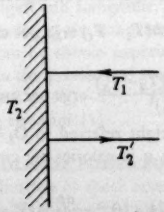

Figure 1. Definition of accommodation coefficient.

If gas molecules at temperature T_1 strike a solid maintained at a different temperature T_2 (see figure 1), the gas molecules leaving the solid will not in general be in temperature equilibrium with it, but their average energy will correspond to some temperature T_2' intermediate between T_2 and T_1 (Knudsen 1911 and 1915).

Provided $T_2 - T_1 \ll T_1$, Knudsen assumed that

$$T_2' - T_1 = a(T_2 - T_1), \qquad \ldots\ldots(1)$$

where a is a constant at a given temperature called the accommodation coefficient. As we shall see later, the correctness of equation (1) can be established by experiment. For polyatomic molecules a will not necessarily have the same value for the rotational and vibrational energy as for the translational energy, and we shall first discuss monatomic gases.

Let us consider the loss of heat from a fine wire stretched down the centre of a cylindrical glass tube which contains a gas at such a low pressure that the mean free-path of the molecules is great compared with the radius of the tube. Let the temperature of the tube be T_1 and let that of the wire be maintained at T_2 by passing an electric current through it. Let the diameter of the wire be much less than that of the tube. Under these conditions each molecule leaving the wire will make many collisions with the tube before returning again to the wire, and it can be assumed that the molecules striking the wire are at the temperature T_1 of the tube. The number of molecules striking unit area of the wire per second is given by

$$\frac{p}{(2\pi m k T_1)^{\frac{1}{2}}}, \qquad \ldots\ldots(2)$$

where p dyne cm^{-2} is the pressure of the gas, m is the mass of a molecule and k is Boltzmann's constant in ergs per degree per molecule. It can be shown that for a monatomic gas the net loss of energy per unit area of the wire per second is given by

$$\frac{p}{(2\pi m k T_1)^{\frac{1}{2}}} \cdot 2k(T_2' - T_1) \text{ ergs per cm}^2 \text{ per sec.}, \qquad \ldots\ldots(3)$$

where the molecules leaving the wire on the average have energy corresponding to the temperature T_2 (see, for example, Roberts, 1939). This result was first deduced by Knudsen. Using equation (1), expression (3) may be written

$$\frac{p}{(2\pi m k T_1)^{\frac{1}{2}}} 2ak(T_2 - T_1) \text{ ergs per cm}^2 \text{ per sec.}, \qquad \ldots\ldots(4)$$

or

$$\frac{7{\cdot}3 \times 10^8 ap(T_2 - T_1)}{(MT_1)^{\frac{1}{2}}} \text{ ergs per cm}^2 \text{ per sec.}, \qquad \ldots\ldots(5)$$

where M is the molecular weight referred to $O_2 = 32$ and p is the pressure in dynes per cm^2. The heat loss q in calories per cm^2 per second is therefore for a monatomic gas given by

$$q = 1{\cdot}74 \times 10^{-4} \frac{ap}{(MT_1)^{\frac{1}{2}}}(T_2 - T_1). \qquad \ldots\ldots(6)$$

By unit area of the wire we mean unit apparent area, and not the actual superficial area. The linear dimensions of irregularities in the surface of the wire will always be small compared with its diameter, and for the long free-paths under consideration the number of molecules striking the wire will not be affected by

these irregularities. They will, however, affect the heat loss and make the measured or apparent accommodation coefficient greater than that for a smooth surface, which we call the true accommodation coefficient, because a certain number of the gas molecules after making a collision with and leaving one part of the surface will strike it again in another place. If the accommodation coefficient for a smooth surface is a_1 and if, owing to roughness, each molecule coming up to the surface on the average strikes the surface in n different places before finally leaving it, it can be shown (Roberts, 1930 a) that the effective value a of the accommodation coefficient is given by

$$a = 1 - (1 - a_1)^n. \qquad \ldots\ldots(7)$$

It is not possible to determine n and thus not possible to deduce a_1 from a. Equation (7) shows that, if n is large, a approaches unity. This fact is used in the work described in § 2.

All the quantities in equation (6) can be measured except a, which can therefore be determined. q is obtained from the electrical energy supplied and the length and diameter of the wire, provided a suitable correction is made for end and radiation losses (Roberts, 1932 a). p can be measured on a MacLeod gauge. T_1 is the temperature of the bath in which the containing tube is immersed, and T_2 is obtained by measuring the resistance of the wire. Actually it is not necessary to work at such low pressure that the mean free-path λ of the gas molecules is large compared with the radius of the containing tube, but it is sufficient that λ should be large compared with the diameter of the wire. On the average, molecules leaving the wire travel a distance λ before making a collision and then, under the conditions specified, make many collisions with other gas molecules before returning to the wire. The passage of the heat from a circle of radius λ to the outside of the tube may be treated as ordinary gas conduction. Simple numerical calculations show that the temperature drop in the gas is small compared with the temperature difference between the wire and the gas at a distance λ from it (see, for example, Blodgett and Langmuir, 1932). The molecules striking the wire can thus be taken as being at the temperature T_1 of the outer tube.

For a given $(T_2 - T_1)$ it can be shown experimentally that at low pressures the heat loss is proportional to the pressure as required by equation (6). For a given pressure it is proportional to $(T_2 - T_1)$. This establishes the correctness of the assumptions underlying equation (1).

§ 2. BEHAVIOUR OF POLYATOMIC GASES

Before discussing the application of these results in connection with experiments on monatomic gases, which is the subject with which we shall be primarily concerned in this part, we shall obtain the analogous equation for polyatomic gases. The average translational energy of a molecule at temperature T is $\frac{3}{2}kT$, and the heat carried away from unit area per second due to changes in the translational energy (see expression (3)), is greater than the number $p(2\pi mkT_1)^{-\frac{1}{2}}$ of

molecules striking unit area per second multiplied by $\frac{3}{2}k(T_2-T_1)$ or by $\frac{3}{2}ak(T_2-T_1)$. The reason for this is that the expression for the number of molecules with velocity between c and $(c+dc)$ striking unit area per second contains a factor c so that, if the numbers per unit volume with velocity between c_1 and c_1+dc_1 and between c_2 and c_2+dc_2 were equal, $c_1>c_2$, more of the faster molecules would strike unit area per second. On the other hand, in considering the internal energy of polyatomic molecules we assume the distribution of rotational and vibrational energy is independent of the distribution of translational energy, i.e. that a fast-moving molecule will not be more likely to have high rotational or vibrational energy than a slow one. If then the average rotational energy at temperature T is $f_r kT/2$ and the average vibrational energy is $f_v kT/2$, the loss of energy per unit area per second is given by

$$\frac{p}{(2\pi mkT_1)^{\frac{1}{2}}}(4a_t+f_ra_r+f_va_v)\frac{k}{2}(T_2-T_1)\text{ ergs cm.}^{-2}\text{ sec.}^{-1}, \quad \ldots\ldots(8)$$

where a_t, a_r, a_v are the accommodation coefficients for translational, rotational and vibrational energy respectively. The heat loss q in calories per cm^2 per second is given by

$$q=4{\cdot}35\times10^{-5}\frac{(4a_t+f_ra_r+f_va_v)p}{(MT_1)^{\frac{1}{2}}}(T_2-T_1). \quad \ldots\ldots(9)$$

From experiments on the heat loss from a fine wire the value of the quantity $(4a_t+f_ra_r+f_va_v)$ can be obtained. Eucken and Bertram (1936) have attempted to measure the molecular heat of a vapour by measuring the heat loss per cm^2 at a low pressure from a rough, oxidized, nickel wire. For such a rough wire it was found that for argon and krypton a was very nearly unity, and it was assumed that $a_t=a_r=a_v=1$, and thus that the experiment gave the quantity

$$(4+f_r+f_v).$$

Now the molecular heat at constant volume at temperature T is

$$(3+f_r+f_v)\frac{R}{2},$$

where R is the gas constant, and thus its value can be deduced if $4+f_r+f_v$ is known. The experiments were carried out with a number of gases or vapours, and the values for the molecular heats were compared with those that would be expected from the analysis of the spectrum, where this was available.

§3. THE EFFECT OF REMOVAL OF ADSORBED FILMS ON THE VALUE OF THE ACCOMMODATION COEFFICIENT

We now return to equation (6) and to the measurement of the accommodation coefficient for monatomic gases. The object of work carried out in this connection must be to obtain a clear and satisfactory picture of the collision process; in other words, to obtain experimental results which can form the basis of a mechanical theory of the collisions. It is obviously necessary for this purpose

to know the nature of the solid and gas atoms involved. This can be seen immediately from the consideration that the expression for the energy transfer at a collision will certainly involve the ratio of the masses of gas and solid atoms. Ordinarily solid surfaces are covered with an adsorbed film of impurities of unknown structure and composition, and it is necessary at the beginning of the experiment to remove all such adsorbed impurities from the surface of the wire and to take such precautions as are necessary to ensure that the gas is free from traces of adsorbable foreign gases so that the surface remains bare. The first experiments satisfying these conditions were carried out with helium and a tungsten wire. In previous experiments by various workers in which an ordinary wire was used, it was found that the accommodation coefficient of helium at room temperature was always about 0·3. With the bare tungsten surface there is a striking change and a value of 0·05 or 0·06 is obtained (Roberts, 1930 b and

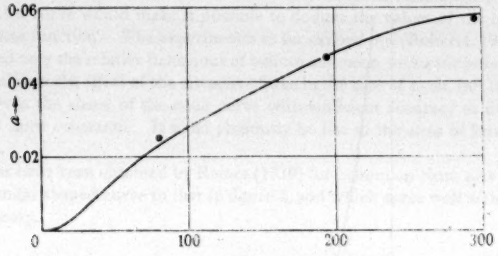

Figure 2. Accommodation coefficient of helium.

1932 a; Mann, 1934; Mann and Newell, 1937). Not only is the actual magnitude of the accommodation coefficient affected, but the nature of its variation with temperature is quite different. Soddy and Berry (1911) found for helium with an ordinary platinum wire a value of 0·28 in the neighbourhood of 20° C. and of 0·36 in the neighbourhood of −185° C. With bare tungsten, on the other hand, the temperature variation is shown in figure 2. The general trend of the points suggests that the accommodation coefficient approaches zero as the temperature approaches zero, but no doubt the lowest portions of the curves would not be realized in practice owing to condensation of the helium.

The reason for this behaviour can be understood from the following simple physical considerations (Roberts, 1932 b). To begin with, let us regard the solid as an assembly of Planck oscillators all of identical frequency ν. When a gas atom interacts with such an assembly, energy exchanges can take place only in amounts of $n\hbar\nu$, where n can have the values 0, 1, 2, 3, etc. At temperatures at which the mean thermal energy is smaller than $\hbar\nu$, a considerable number of the oscillators will be in the ground state. When a gas atom interacts with an oscillator in the

ground state, the only possible interchange of energy is that in which the gas atom gives up energy to the oscillator, and the smallest amount of energy that an oscillator can take is $h\nu$. The number of gas atoms that have energy $h\nu$ to give up becomes progressively smaller as the temperature gets lower, and thus the proportion of gas atoms which can undergo a change of energy on interacting with the solid becomes smaller and smaller as the temperature approaches the absolute zero. If the solid is not regarded as an assembly of oscillators all having identical frequencies, but as having frequencies distributed over a range 0 to ν_m, the above considerations show that at sufficiently low temperatures only the low-frequency oscillations will be appreciably affected by collisions with gas atoms. Therefore the number of oscillations so affected becomes fewer and fewer as the temperature becomes progressively lower, and as a result the interchange of energy, other

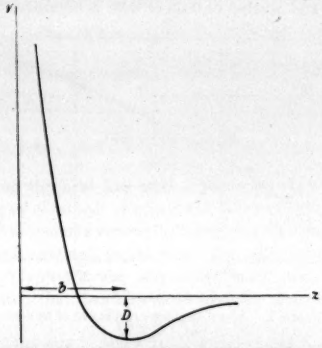

Figure 3. Potential energy V of two molecules at distance z apart, showing meaning of D and b in the Morse function.

things being equal, becomes less and less efficient. Thus for a small temperature difference between solid and gas the accommodation coefficient becomes smaller and smaller.

Devonshire (1937), in one of a series of papers by Lennard-Jones and his co-workers, has followed up the work of Fowler and of Jackson and Mott and others (Jackson, 1932 ; Jackson and Mott, 1932 ; Jackson and Howarth, 1933, 1935 ; Landau, 1935 ; Michels, 1937) in developing the detailed quantum-mechanical theory of collisions of gas atoms with solids. The efficiency of energy interchange between the gas atom and the oscillations of the solid depends among other things on the mass of gas and solid atoms, which are known, and on the law of force between the gas atom and the atom of the solid. Devonshire assumes that the potential energy when their centres are at a distance z apart is given by a Morse function

$$De^{-2\kappa(z-b)} - 2De^{-\kappa(z-b)}. \qquad \ldots\ldots(10)$$

This is plotted in figure 3, which shows the meanings of b and of D, which is closely related to the heat of adsorption. The further constant κ determines the rate at which the repulsion potential energy rises when the atoms are closer together than b. The greater κ is, the sharper is this rise and, other things being equal, the greater is the accommodation coefficient. For helium D is small, and the curve in figure 2 is Devonshire's theoretical curve for a one-dimensional model for $\kappa = 2 \times 10^8$ and $D = 0$, multiplied throughout by a roughness factor 1·06, which is chosen to make the curve pass through the experimental point at 195° K. The theory of course gives a_1, the accommodation coefficient for a smooth surface, and equation (7) shows that for small values of a_1 the effective accommodation coefficient is equal to na_1, where n is a constant for a given surface.

For neon, D will certainly be appreciable. This makes the curve showing the variation of a with temperature fall off less slowly than it would if D were zero, and Devonshire has shown that an accurate experimental determination of the shape of this curve would make it possible to deduce the values of the constants in the Morse function. The experiments so far carried out (Roberts, 1933) have determined only the relative behaviour of helium and neon with a tungsten surface and have shown the effect of the attractive force in the case of neon, but they have not yet given the shape of the neon curve with sufficient accuracy to determine the actual force constants. It must obviously be one of the aims of future work to do this.

Results have been obtained by Raines (1939) for helium on clean nickel which show a similar shaped curve to that in figure 2, and which agree well with Devonshire's theory.

§ 4. THE STUDY OF ADSORPTION PROCESSES USING THE CHANGE OF ACCOMMODATION COEFFICIENT AS AN INDICATOR

With neon the difference at room temperature between the accommodation coefficient for an ordinary surface covered with adsorbed films (about 0·6) and for a clean tungsten surface (about 0·06) is very large. This suggests that the accommodation coefficient of neon is very sensitive to the presence of adsorbed films and that therefore it can be used satisfactorily as an indicator in studying the adsorption of gases on bare tungsten.

The essential parts of apparatus suitable for the application of this method to the study of the adsorption of hydrogen (Roberts, 1935 a) are shown diagramatically in figure 4. The tungsten wire was contained in the tube A. The neon was continuously circulated by a diffusion pump through the charcoal tubes B immersed in liquid air to remove from it any adsorbable impurities coming from the glass. There were liquid-air traps between the charcoal tubes and the rest of the apparatus, and two such traps between the tube containing the wire and the MacLeod gauge.

After the neon had been circulated for some time, the wire, which had previously been thoroughly outgassed, was heated to a temperature well above

2000° K., so as to remove all adsorbed impurities from its surface. It was connected to a Wheatstone bridge and, when it had cooled down, a measured current was passed through it sufficient to raise its temperature about twenty degrees above that of the oil bath in which the containing tube was immersed. Taking the time of starting the experiment as the moment when the flashing current was cut off, readings of resistance as a function of the time were taken and the pressure was read on the MacLeod gauge. From each reading a value of the accommodation coefficient was deduced. These values are plotted on the first part of the curves of figure 5, which refer to experiments at 295° K. and 79° K. They show

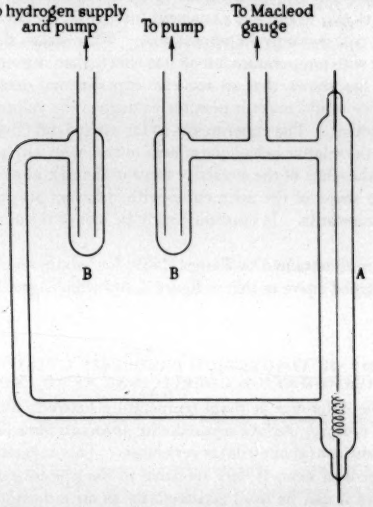

Figure 4. Apparatus for measuring accommodation coefficient.

a very slow drift with time, presumably due to the adsorption on the wire of very small residual traces of impurity in the neon. Extrapolation to zero time gives the accommodation coefficients for a clean wire (see Roberts, 1935 a, footnote on p. 447; Michels, 1937). At the time indicated by the arrow a dose of hydrogen contained between two taps forming a gas pipette, and sufficient to produce a pressure of about 10^{-4} mm. mercury in the apparatus, was admitted so that it had to pass through a charcoal tube before reaching the wire. The accommodation coefficient rose rapidly to a final steady value. This indicates adsorption of the hydrogen on the tungsten. The final steady value was not affected if very much larger doses of hydrogen were admitted, which suggests that a practically complete

monomolecular layer is formed even when the hydrogen is present at a partial pressure of only 10^{-4} mm. of mercury, and this view is confirmed by the experiments of a different kind described in § 5.

Blodgett and Langmuir (1932) had previously measured the accommodation coefficient of hydrogen with a tungsten surface. Their experiments showed certain changes in the accommodation coefficient according to the heat treatment of the wire, and they interpreted these as due to changes in an adsorbed film of hydrogen on the surface. The experiments just described show conclusively that their view that a hvdrogen film is formed on a bare tungsten surface is correct.

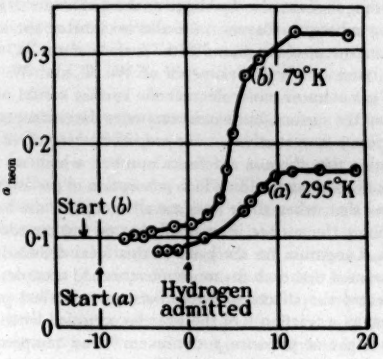

Figure 5. Accommodation coefficient of neon.

For gases like oxygen and nitrogen, which are strongly adsorbed by charcoal, some modification of the apparatus is necessary. In fact, subsidiary experiments with the hydrogen apparatus to prove that the effects observed with hydrogen were not due to traces of air, which might have been present as an impurity, showed that, when the dose of hydrogen was replaced by an equivalent dose of air, no observable effect was produced. For such gases the apparatus in its latest form is discussed in § 7. With this apparatus it is possible to obtain a known steady and extremely low partial pressure of, say, oxygen and, using it, to study in great detail, as we shall see, the structure of the adsorbed film of this gas.

While up to the present the number of systems that has been studied by these methods is limited, yet the results obtained are very definite, and the viewpoint as to the structure of adsorbed films based on the ideas they have suggested must be regarded as of wide applicability.

PART 2. THE STRUCTURE OF ADSORBED FILMS

§ 5. MONOMOLECULAR LAYERS

A great advance was made in the theory of adsorption when Langmuir introduced and developed the idea of monomolecular layers. Most of the work previous to this had been carried out with substances like charcoal, the effective superficial area of which was quite unknown. Langmuir measured the adsorption on a large number of sheets of mica of known area and showed that at saturation the amount adsorbed was of the same order as the amount that could be packed into a complete layer on the surface of the mica, and this led him to the view that the forces holding the first adsorbed layer on the surface are in general stronger than those holding subsequent layers. He also introduced the idea that, owing to the regular arrangement of the atoms in the surface of a solid, which had been recently demonstrated by the experiments of W. H. and W. L. Bragg, the force acting on a gas atom or molecule near the surface would not depend only on its distance from the surface, but would vary over the surface so as to be much stronger at some points than at others. He proposed taking these two effects into account by assuming that there is a definite number, which we shall call n_s, of sites per unit area of solid surface on which adsorption of gas atoms or molecules can take place, and that, when these sites are all occupied, the force acting on a molecule approaching the surface is so small that no further adsorption occurs. As is well known, Langmuir, on the basis of this idea, deduced his adsorption isotherm. He assumed that each gas molecule adsorbed occupies one site on the surface and neglected the effects of forces between adsorbed particles. With these assumptions, let a fraction θ of the sites be occupied when equilibrium is established with a gas at pressure p dynes cm^{-2} and temperature T. The number of molecules striking unit area of the solid per second is $p(2\pi mkT)^{-\frac{1}{2}}$. If a fraction α of those striking vacant sites condenses, the number condensing in unit area per second is $\alpha(1-\theta)p(2\pi mkT)^{-\frac{1}{2}}$. With the assumption of no interaction between adsorbed particles, the number evaporating per second is proportional to the number present, i.e. to θ, and is put equal to $A\theta$ per unit area, where A is a function of the temperature. For equilibrium the rates of condensation and evaporation are equal. Thus

$$\frac{\theta}{1-\theta} = Bp, \qquad \ldots\ldots(11)$$

where B is a function of the temperature. Before we discuss further the assumptions on which this equation is based, we shall consider some more recent experiments which support the idea of the monomolecular nature of adsorbed layers.

Palmer and Clarke (1935) have measured the adsorption of various vapours on powdered fused silica, the surface area of which was known. They plotted p against p/x, where x is the amount adsorbed at pressure p. At low surface concentrations for acetone, for example, the relation was a straight line, as it should be if a Langmuir isotherm is followed. From the slope of this line they obtained the amount x_s in a complete monomolecular layer and showed that this

agreed with the amount that could be packed on to the surface as calculated from the cross-sectional area of the molecules, and pointed out that in cases like this it is not necessary to assume definite sites for adsorption to obtain a Langmuir isotherm. The work has later been extended to higher pressures, and the results show that the ordinary saturation vapour-pressure is reached when four layers are complete (Palmer, 1937 and 1938). This may be compared with the results discussed in § 9.

Oliphant (1928) has carried out some very ingenious experiments on the adsorption of carbon dioxide on drops of falling mercury. This substance has the great advantage that the geometrical or apparent surface area is the true surface area, and no unknown factor need be introduced to take account of roughness. He has shown that about 0·6 or $0{\cdot}7 \times 10^{15}$ molecules are adsorbed per cm^2. The gas-kinetic diameter of CO_2 is $4{\cdot}2 \times 10^{-8}$ cm., and, if the molecules on the surface occupy a square with a side of this length, a monomolecular layer on one cm^2 would contain approximately this number of molecules.

More recently Burdon (1935) has measured the amount of hydrogen adsorbed on mercury by forming a new surface of the metal in the presence of hydrogen at a pressure of 400 mm. of mercury. The film was held tenaciously when the pressure was reduced to 2×10^{-4} mm., and the amount in it was determined by withdrawing the mercury into a much smaller tube so that its surface area was reduced. The adsorbed hydrogen was given off and its amount estimated by measuring the pressure developed. It was found that it was of the same order as would correspond to a complete monomolecular layer on the surface.

Finally, we shall consider some experiments (Roberts, 1935 a) in which the amount of hydrogen adsorbed on a fine tungsten filament 0·0066 cm. in diameter and 28·2 cm. long was measured. The experiments described in § 4 indicate that a complete film of hydrogen is formed rapidly on the tungsten and that it is stable at very low pressures. A simple calculation shows that the amount of gas in a monomolecular film on a wire of the above size would exert a measurable pressure if it occupied a volume of, say, 50 c.c. By admitting successive equal doses of hydrogen to an apparatus of which the known volume was of this order and which contained an initially bare tungsten filament, it was possible to measure the amount of hydrogen adsorbed on the tungsten. Each dose was less than sufficient to cover the surface completely, and the amount in a dose was determined by measuring with a Pirani gauge the pressure change on admission after the wire was covered. The first doses admitted were completely adsorbed and produced no measurable change of pressure, i.e. the change was less than 2 or 3×10^{-7} mm. The total numbers of molecules adsorbed in five independent experiments were respectively 4·2, 4·4, 4·3, 3·3 and $3{\cdot}8 \times 10^{14}$. We take the largest, $4{\cdot}4 \times 10^{14}$, as being nearest to the correct value and assume that the lower values in the other experiments are due to a slight contamination of the wire before the admission of the hydrogen. The effective apparent superficial area of the wire was 0·55 cm^2 after correcting for short lengths at the ends, which during the initial cleaning

would not have been heated to a sufficiently high temperature to remove any adsorbed oxygen, and would not therefore be available for the adsorption of hydrogen.

We shall compare the number of molecules adsorbed with the number of tungsten atoms in the surface. It has been usual, following Langmuir (1932), to assume that with aged tungsten the (110) plane is the important one in the surface, but R. P. Johnson (1938) has shown that this is not necessarily correct and that the (100) plane must also be considered. Tungsten is a body-centred cubic lattice, and the arrangements of the atoms in these two planes are shown in figure 6. The numbers of atoms per cm^2 are for the (110) plane $14{\cdot}24 \times 10^{14}$ and for the (100) plane $10{\cdot}07 \times 10^{14}$. Thus, if the wire were smooth, the number of tungsten atoms in the surface would be $7{\cdot}8 \times 10^{14}$ for the (110) plane and $5{\cdot}5 \times 10^{14}$ for the (100) plane. If the surface were not smooth these numbers would have to be multiplied by a factor ρ greater than unity. The high stability of the adsorbed film suggests that we have to do with chemisorption, which can be presumed to

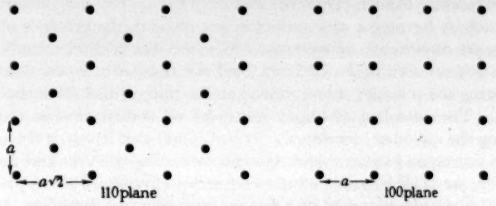

Figure 6. Atomic arrangement in surface planes of tungsten.

involve a dissociation of the hydrogen molecule and the adsorption of the two atoms separately.* If there were one atom of hydrogen for each tungsten atom in the surface, the number of adsorbed molecules would be $\rho \times 3{\cdot}9 \times 10^{14}$ for the (110) plane and $\rho \times 2{\cdot}75 \times 10^{14}$ for the (100) plane. A reasonable value of ρ can be taken as somewhere between 1 and 2. Thus, whichever plane predominates, the measured amount of adsorbed hydrogen, $4{\cdot}4 \times 10^{14}$ molecules, is in agreement with the view of the monomolecular nature of the film.

In these experiments the heat of adsorption of the hydrogen was also measured. The wire formed one arm of a sensitive Wheatstone bridge in which a very small measuring current was used. The heat of adsorption was obtained from its thermal capacity (deduced from its dimensions, density and specific heat) and the rise of temperature when the hydrogen was adsorbed. The results are shown in figure 7, in which the heat of adsorption is plotted for five different experiments as a function of n_f, the number of filled planes on the surface. A different sign is used for plotting each experiment, and it may be mentioned that the relative

* See Roberts (1939), p. 81. A full discussion of evaporation processes, including the production of atomic hydrogen, is given in chapter IV of the Tract referred to.

values of the heats for the successive admissions in a particular experiment can be fixed with considerably higher accuracy than the absolute values which are used in comparing different experiments. If we extrapolate linearly we obtain, for $n_f = 0$ a heat of adsorption of 45 000 calories per mol H_2 and of 18 000 calories when $n_f = 8{\cdot}8 \times 10^{14}$, which corresponds to a complete film.

These large changes of heat as the surface becomes covered are evidence of the existence of strong repulsive forces between the adsorbed atoms, and it therefore becomes a matter of immediate importance to examine what modifications

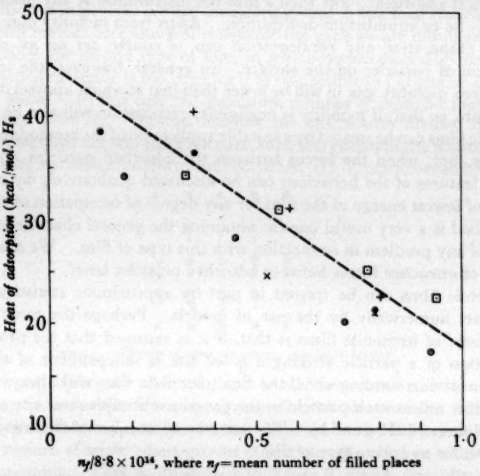

Figure 7. Relation between heat of adsorption and amount adsorbed.

of the results of the simple theory, which neglects the effects of such forces, are caused by taking them into account. A little reflection shows that the development of the theory including these effects makes it necessary to consider in detail the structure of adsorbed films and to re-examine carefully the whole basis of the theory. We shall now give a summary of the necessary developments from the physical point of view, and shall consider only those cases which have a direct application to experimental results.

§6. MOBILE AND IMMOBILE ADSORBED FILMS

Adsorbed films on solids may be divided into two general classes, mobile and immobile films. By a mobile film we mean one in which the energy of activation necessary to enable an adsorbed particle to move from a site on which it is adsorbed is much less than kT, so that the particles move freely from one site to another.

This free movement enables the film to take up equilibrium configurations during the occurrence of any process. Thus for a mobile film at any instant the particles at which free mobility sets in will be lower than that at which appreciable evaporation occurs, so that, if mobility is negligible, evaporation will also be negligible.

For an immobile film, on the other hand, the energy of activation necessary to enable a particle to move from one site to another is so much greater than kT that, for the times with which we are concerned in any given experimental investigation, the particles may be treated as remaining fixed on the sites in which they are first adsorbed. For such a film the distribution at any instant will not necessarily be an equilibrium distribution. Apart from mobility of the adsorbed particles, evaporation and condensation can, of course, set up an equilibrium distribution of particles on the surface. In general, however, the temperature at which free mobility sets in will be lower than that at which appreciable evaporation occurs, so that, if mobility is negligible, evaporation will also be negligible.

Mobile films can be treated by a suitable application of the methods of statistical mechanics, but, when the forces between the adsorbed particles are *repulsive*, the main features of the behaviour can be discussed qualitatively by considering the state of lowest energy of the film for any degree of occupation of the surface. This method is a very useful one for obtaining the general characteristics of the solution of any problem in connection with this type of film. We shall consider the effect of *attractive* forces between adsorbed particles later.

Immobile films can be treated in part by approximate statistical methods and in part numerically by the use of models. Perhaps the most important characteristic of immobile films is that, if it is assumed that the probability of condensation of a particle striking a given site is independent of the state of occupation of surrounding sites, the final complete film will always have gaps or holes in it unless each particle in the gas phase occupies one site only on the surface (Roberts, 1935 a and b). The nature and number of these gaps must be determined for any given type of film.

§7. GAPS IN IMMOBILE FILMS AND THE STRUCTURE OF THE OXYGEN FILM ON TUNGSTEN

It is known from experiments on the thermionic emission from tungsten that an oxygen film, which can only be removed by heating to temperatures in the neighbourhood of 1700° to 2000° K., is adsorbed on this metal. Since this film is so stable to evaporation it seems reasonable to suppose that at temperatures as low as room temperature the film will be immobile and that a molecule once adsorbed will remain fixed in position. The diameter of the oxygen molecule is greater than the spacing of the tungsten atoms in the surface. Thus one oxygen molecule cannot be accommodated on each site for adsorption. We can suppose either that the oxygen dissociates on adsorption and that the two atoms are adsorbed on neighbouring sites, or that the oxygen is adsorbed as molecules and that, when a molecule is adsorbed on a given site, adsorption cannot take place on the four sites surrounding this site, i.e. that the nearest site to an occupied

site on which adsorption of another molecule can take place is one of the next nearest neighbours. In the simple quadratic lattice shown in figure 8, this means that, if site 0 is occupied, adsorption can occur on sites 5, 6, 7, 8, but not on sites 1, 2, 3, 4.

Let us consider the first type of film, in which the molecules dissociate on adsorption and the atoms are adsorbed on neighbouring sites. It is evident that, as the film is built up, certain single isolated sites will be unoccupied while all the neighbouring sites will be occupied. These single unoccupied sites can take no part in the adsorption process we are considering, and in the final complete film of this type will remain bare and will constitute the gaps we are discussing.

Now consider the second type of film, in which dissociation does not occur but each adsorbed molecule is too large for it to be possible to put one on each site as described above. In this case the maximum number of particles that can be put on N_s sites is $0{\cdot}5\,N_s$. We can imagine the lattice sites divided into two groups, which we shall call odd and even sites, such that each odd site is surrounded by four even sites and each even site by four odd sites. $0{\cdot}5\,N_s$ particles can be packed

Figure 8. Group of sites on simple quadratic lattice.

on the surface only if they are all on odd sites or all on even sites. In the process of building up an immobile film it is obvious that all the particles will not be adsorbed on sites of one type only, and that therefore in the final complete film there will be regions of misfit where groups of particles adsorbed on odd and even sites are contiguous. This is illustrated in figure 9, which was obtained by occupying at random a square arrangement of 100 numbered sites by drawing cards. The odd and even sites are given odd and even numbers, and to avoid edge effects the sites are imagined to be arranged in circles around a cylinder the ends of which are joined together. This is represented on the plane diagram by repeating the outer sites as shown in figure 9.

The regions of misfit, or boundary sites, constitute the gaps in this case. If we call a site which is surrounded by four unoccupied sites a first-layer site, a boundary site may be defined as an unoccupied site which is not surrounded by four occupied or by four unoccupied sites and which has no first-layer site as a nearest neighbour.*

* This is the definition given by the author (Roberts, 1940 a). It was not stated in that place (but implied) that the vacant site must not be surrounded by four occupied sites. For the discussion in that paper this was the most convenient definition. A definition that is equivalent when the film is complete, but which could give different numbers of boundary sites in an incomplete film, is the following :—A site with at least one neighbouring site occupied and at least one neighbouring site vacant, but unable to be occupied.

To whichever type the oxygen film on tungsten belongs, the effect is qualitatively the same. We have the tightly held film with gaps in it. It seems likely that over these gaps the forces tending to hold oxygen molecules will be considerably stronger than at other points of the surface where first-layer film is packed as closely as possible.* Bearing this in mind, let us consider the experimental results obtained by Morrison and the author (1939). As before, the accommodation coefficient of neon was used as an indicator of the presence of the adsorbed film, and the apparatus used is shown in figure 10. The object of this

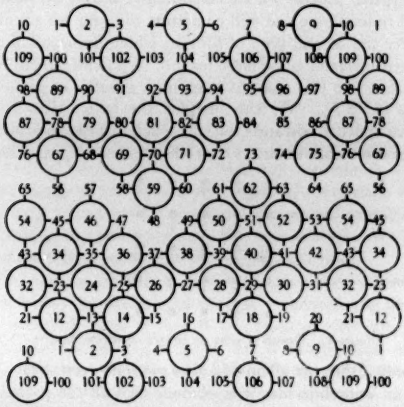

Figure 9. Model of simple quadratic lattice occupied by random selection of sites. The numbers represent sites and the circles adsorbed particles. Adsorption of a particle on a given site makes it impossible, as indicated by the short lines, for adsorptions to occur on the four neighbouring sites.

particular form of apparatus was to make it possible to determine the partial pressure of the oxygen at the wire W on which adsorption took place, and thus to make a quantitative study of the phenomena. The whole apparatus was filled with neon at a pressure of the order of 0·1 mm. C_1 and C_2 were the charcoal tubes immersed in liquid air through which the neon was continuously circulated. After the wire had been heated to above 2000° K. to remove all adsorbed impurities from its surface and had cooled down, a measured current was passed through it sufficient to raise its temperature about twenty degrees above that of the bulb in which it was immersed, and the accommodation coefficient of the neon was obtained as before.

* On the whole the quantitative evidence is against the first-layer film being atomic, but this evidence cannot be regarded as conclusive (Roberts, 1940 a).

At a given instant a measured dose of oxygen was admitted from a gas pipette to the calibrated 2-litre bulb B, so that the pressure of oxygen in B was known. This oxygen diffused through the long fine capillary-tube F. On reaching the far end it divided into three streams. The first filled the MacLeod gauge and, once a steady state had been reached, ceased. Another went to the charcoal tube C_1, where it was removed. The third went past the wire W mostly to C, which was immersed in liquid air, and a very small fraction to C_2.

We require to obtain the pressure of oxygen at W from the known pressure in the bulb B. The resistance to the flow of the various streams of oxygen is due in part to the effect of the walls of the tubes and in part to the fact that it is diffusing through neon. The apparatus was so designed that in the tube F the resistance due to the walls was greater than that due to collisions with neon

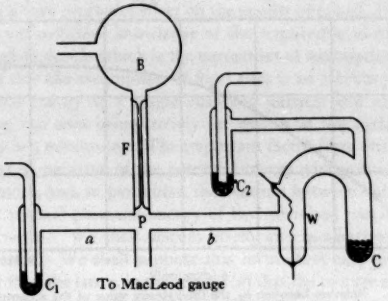

Figure 10. Modified apparatus.

atoms, so that the latter was only a correction term. In the part of the apparatus from P to C_1 and C, on the other hand, most of the resistance was due to collisions with neon atoms, and the effect of the walls was only a correction term. These resistances could all be calculated from the dimensions of the tubes, the pressure of the neon, and kinetic-theory data, including the diffusion coefficient. The results of the calculation may be expressed in the form familiar in similar electrical problems by writing the flow through a number of tubes in series as equal to the pressure drop divided by the resistance of the tubes in appropriate units. Once the MacLeod gauge is filled we may equate the flow through F to the sum of the flows from P to C_1 and P to C (neglecting for the sake of simplicity the very small flow to C_2). Thus we obtain, if p, with an appropriate subscript, is the partial pressure of the oxygen,

$$\frac{p_B - p_P}{R_{PB}} = \frac{p_P}{R_{PC_1}} + \frac{p_P}{R_{PW} + R_{WC}}.$$

Actually p_P was very small compared with p_B, so that we may write

$$\frac{p_P}{} \ldots \left(\frac{1}{} \quad \frac{1}{} \right).$$

R_{PC}, and R_{WC} included very small correction terms due to the " resistance " at the surface of the charcoal, which were determined in subsidiary experiments. This equation gives us p_P in terms of p_B. We have, further,

$$p_W/p_P = R_{WC}/R_{PC},$$

which gives p_W in terms of p_P and thus in terms of p_B. The ratio of p_W to p_B depended markedly on the pressure of the neon, and it was shown that, when this pressure was varied over a wide range, consistent results were obtained.

The accommodation coefficient of neon with bare tungsten is 0·057 at about 300° K., the temperature at which the experiments were carried out. When oxygen is admitted to B, this rises to a final steady value a_f which depends on the pressure of oxygen at the wire. a_f is plotted against the oxygen pressure in figure 11. At very low partial pressures of oxygen the value $a_f = 0{\cdot}226$ is reached.

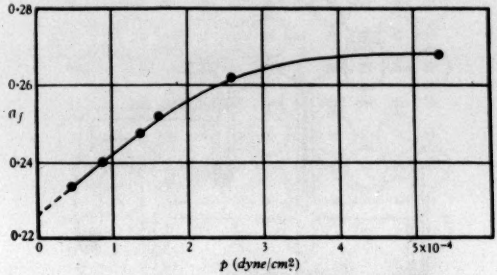

Figure 11. Relation between a_f, the final steady value of the accommodation coefficient of neon, and p, the partial pressure of oxygen.

This result suggests the presence of a very stable film giving $a = 0{\cdot}226$ and a less stable film on top of it, the amount in the latter depending on the oxygen pressure. Further investigation shows that the film corresponding to $a = 0{\cdot}226$ is itself composite, for heating it to about 1100° K. reduces the accommodation coefficient to 0·177, while no further evaporation takes place until the film is heated to above 1700° K.

These results can be interpreted in a natural way in terms of the picture put forward at the beginning of this section. The state $a = 0{\cdot}177$ must be presumed to correspond to the stable immobile film with the gaps in it unfilled. The state $a = 0{\cdot}226$, reached at very low pressures of oxygen, corresponds to the stable film with the gaps occupied by rather more loosely held molecules. The higher values of a shown in figure 11 are due to the presence on top of this structure of a film the equilibrium amount in which depends on the pressure.

It may be noted that, since 1 dyne cm.$^{-2}$ is of the order of 10^{-3} mm., the pressures at which this last film is formed are so low that the forces holding it must be presumed to be very strong. The very low pressures which can be obtained and

measured by the method adopted may be noted. The absolute pressures are correct to about 10 or 20 per cent, but their relative values are much more accurate than this, as is indicated by the accuracy with which the observed points in figure 11 fall on a smooth curve. The existence of the upper layer was confirmed by a study of the kinetics of the formation of the film.

§ 8. EFFECT OF REPULSIVE FORCES BETWEEN ADSORBED PARTICLES FOR FILMS IN A SIMPLE QUADRATIC LATTICE WITH FIXED SITES

We saw in § 5 that the large and progressive change in the heat of adsorption of hydrogen on tungsten as the surface becomes covered suggests the existence in this system of strong repulsive forces between the adsorbed particles. In discussing the effect of such forces from the theoretical point of view, the assumptions made have a very profound effect on the results obtained, and unfortunately we have not as yet sufficient knowledge of the required kind about the various factors concerned to decide which is the correct set of assumptions.

It is evident that the assumption of fixed sites is an arbitrary one and that, in fact, the potential energy of a single adsorbed particle will vary continuously over the surface, the sites being merely the points on the surface at which the potential energy is a minimum. The important factor from our present point of view * is the rate of variation of the potential energy with position as we go away from the minimum, and, in particular, the relation between this and the space-gradient of the mutual potential-energy of two adsorbed particles. Neither of these rates is known. We shall discuss briefly and qualitatively two important cases that can arise. We shall suppose that in the first case (*a*) the former rate is much greater than the latter, in the second (*b*) that the two are of the same order. A limiting example of case (*b*) is obtained when the rate of variation with their distance apart of the mutual potential energy of the particles is much greater than the rate of variation of the potential energy of a single particle with its position. This limiting case need not be discussed in detail, as the behaviour can be inferred from that in (*b*) (see Miller and Roberts, 1941).

Case (*a*). If the rate of variation with position near the minimum of the potential energy of an adsorbed particle is much greater than the rate of variation with the distance apart of the potential energy of the adsorbed particles, a model of the film which will represent the actual behaviour closely is obtained by supposing that the sites are fixed and that there is a fixed interaction energy V between particles adsorbed on neighbouring sites. We suppose, further, that the mutual repulsive forces fall off sufficiently rapidly for the potential energy, due to these forces, of particles adsorbed at a greater distance apart than on neighbouring sites to be negligible.

We shall suppose that the sites are arranged on a simple quadratic lattice, as shown by the intersections of the lines in figure 12. Let us first consider a mobile film and represent the adsorbed particles by black dots. It will be seen

* Cf. Roberts, 1939, p. 38.

that, with the arrangement shown, half the sites are occupied and no two neighbouring sites are occupied, i.e. $\theta = 0{\cdot}5$ and the interaction energy is zero. This is the state of lowest energy for $\theta = 0{\cdot}5$, and it is the state in which the mutual repulsive forces between the molecules will tend to make them arrange themselves. In accordance with Boltzmann's law, the thermal motion of the particles will make it possible for configurations of higher energy to occur, and the average potential energy of the system, due to the repulsive forces, will be greater than zero. If we neglect the effect of the thermal motion and assume that the system is always in the state of lowest energy, it will be seen that up to $\theta = 0{\cdot}5$ the heat of adsorption will remain constant at, say, q_0 calories per particle. It will then fall and remain

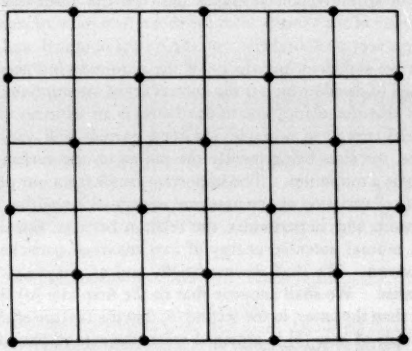

Figure 12. State of lowest energy for adsorbed film on quadratic lattice with repulsion between particles adsorbed on neighbouring sites.

constant at $(q_0 - 4V)$ from $\theta = 0{\cdot}5$ to $\theta = 1$. If the heat is plotted against θ, the curve will have a sharp step at $\theta = 0{\cdot}5$, as shown in figure 13, curve (a) The effect of the Boltzmann distribution will be to round off the sharp corners to a greater or less extent, and at very high temperatures, where $V/kT \to 0$, the distribution of particles on the surface will, for all values of θ, be a random distribution, i.e., the tendency for the particles to arrange themselves in the state of lowest energy will be completely overcome by the thermal motion, and the relation between heat of adsorption and θ will be linear, as shown in curve (e) in figure 13. The actual rounded curves for some intermediate cases specified by the value of the quantity $\eta = e^{-V/kT}$ are shown in curves (b), (c), (d). The details of the calculations of these curves, which were carried out using Bethe's approximation, need not be discussed here (see Roberts, 1939, pp. 31–4, where references are given).

In the kinetic derivation of the Langmuir isotherm, the equilibrium state is seen to be due to the balance between evaporation and condensation of particles

in the film. Returning to a film which is in the state of lowest energy, it can be seen that any particle put in one of the vacant places in figure 9 will require less energy to remove it than one of the other particles shown in the diagram. That is, the rate of evaporation of particles surrounded by four occupied sites will be greater than that of particles surrounded by four vacant sites. Thus we should expect a considerable change in the slope of the curve showing the relation between θ and p in the neighbourhood of $\theta = 0{\cdot}5$, such that the change in p for a given change in θ is greater above $\theta = 0{\cdot}5$ than it is below $\theta = 0{\cdot}5$.

If, on the other hand, the film is immobile, the behaviour is different, for in this case, if a molecule strikes a vacant site, the fraction of the surrounding sites

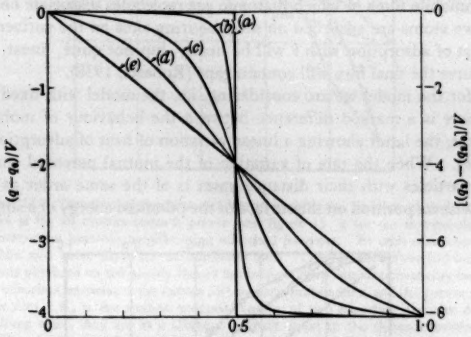

Figure 13. Variation of heat of adsorption with fraction of surface covered.
(*a*) $\eta = e^{-V/kT} = 0$; (*b*) $\eta = 3{\cdot}6 \times 10^{-3}$; (*c*) $\eta = 8{\cdot}2 \times 10^{-2}$; (*d*) $\eta = 0{\cdot}368$; (*e*) $\eta = 1$. The scale on the left gives q for no dissociation and that on the right gives q_2 when there is dissociation. The subscript 0 refers to $\theta = 0$, and V is the interaction energy between particles adsorbed on neighbouring sites. The calculations are for $z = 4$.

that is occupied is proportional to θ. Thus the variation of heat of adsorption with θ is linear from q_0 at $\theta = 0$ to $(q_0 - 4V)$ at $\theta = 1$. The final film in this case is complete and contains no gaps, since we are assuming that each particle in the gas phase occupies one site on the surface.

For mobile films in which diatomic gas molecules dissociate on adsorption and in which the two atoms are adsorbed on neighbouring sites on the surface, the heat of adsorption q_2 of a molecule is defined as the decrease in the energy of the system when a molecule is adsorbed. If q_1 is the heat of adsorption of an atom, it is easy to show from a simple cycle of changes that

$$q_2 = 2q_1 - q_d.$$

Thus, if $(q_2)_0$ is the heat of adsorption of a molecule when $\theta = 0$ and V is the interaction energy between two *atoms* adsorbed on neighbouring sites, the curves of

figure 13 give $q_2-(q_2)_0$, provided the scale on the right-hand side of the figure is used.

If the system is in the state of lowest energy, the atoms will arrange themselves on the surface as shown in figure 12, and when $\theta=0{\cdot}5$ no further adsorption can take place because at no place will there be two neighbouring vacant sites on the surface. Thus, whenever the repulsive interaction between adsorbed particles is appreciable, we should expect a marked falling-off in the rate of adsorption in the neighbourhood of $\theta=0{\cdot}5$ and, as in the case when each particle in the gas phase occupies only one site, the isotherm will show a marked change of slope in the neighbourhood of $\theta=0{\cdot}5$.

For immobile films in which diatomic gas molecules dissociate on adsorption and the two atoms are adsorbed on neighbouring sites on the surface, the variation of heat of adsorption with θ will be nearly, but not quite, linear, and in this case of course the final film will contain gaps (Roberts, 1938).

Thus for the model we are considering, i.e. the model with fixed interaction energy, there is a marked difference between the behaviour of mobile and immobile films, the latter showing a linear variation of heat of adsorption with θ.

Case (*b*). When the rate of variation of the mutual potential energy of two adsorbed particles with their distance apart is of the same order as the rate of variation with its position on the surface of the potential energy of a single particle,

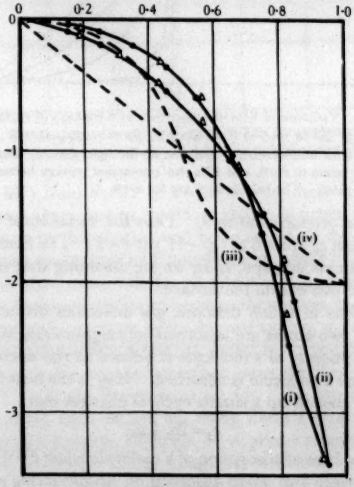

Figure 14.

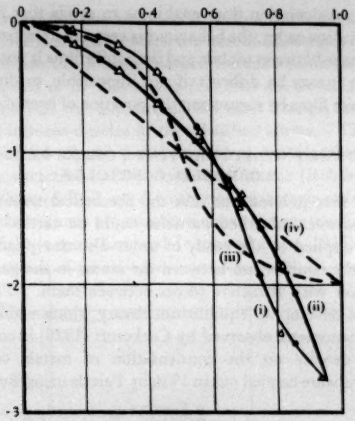

Figure 15.

Figures 14 and 15. Variation of the heat of adsorption with the fraction of the surface covered. Figure 14 is for an inverse-seventh power and figure 15 is for an inverse-eleventh power law of interaction between neighbouring adsorbed particles. In each case, curve (i) is for a mobile film and curve (ii) is for an immobile film. The broken curves (iii) and (iv) are the heat curves obtained on the simple theory assuming a fixed interaction-energy between neighbouring adsorbed particles ; the curves (iii) are for a mobile film, and the curves (iv) are for an immobile film. V_a is the mutual potential energy of two adsorbed particles due to the repulsive force when they are at a distance a apart equal to the distance between two sites. Curves (i) and (iii) are calculated for $\eta=0{\cdot}1$, while curves (ii) and (iv) are independent of η, which is $e^{-V_a/kT}$.

Miller and the author (1941) have shown, by considering a one-dimensional model, that the behaviour is very different from that in case (*a*).

An arbitrary model was used, in which it was assumed that the potential energy of a single particle varied with its position x in the one-dimensional film according to the law

$$-\chi+X(1-\cos 2\pi x/a), \qquad \ldots\ldots(12)$$

and that the mutual potential energy of two adsorbed particles in neighbouring potential hollows varied with their distance y apart according to an inverse power law Yy^{-m}. The "sites" for adsorption are the lowest points of the potential hollows given by expression (12) and are a distance a apart. The results are plotted in figures 14 and 15.

It must be emphasized that this calculation does not necessarily give the behaviour that would be observed if the potential energy of two adsorbed particles due to their mutual repulsive forces obeyed inverse-seventh and inverse-eleventh power laws, because the actual behaviour depends not only on this law, but also on the way the potential energy of a single particle varies with its position on the

surface. What the calculation does enable us to say is that if the two rates of variation are of the same order, the behaviour is very different from that in case (*a*), and that the difference between mobile and immobile films is not nearly so marked. This suggests that it may be difficult, if not impossible, to distinguish between mobile and immobile films by measuring the variation of heat of adsorption with θ.

§9. EFFECT OF ATTRACTIVE FORCES BETWEEN ADSORBED PARTICLES

Fowler (1936) first pointed out that the theoretical treatment of problems involving forces between adsorbed particles could be carried out using similar methods to those applied to the study of order-disorder phenomena in alloys, and he discussed the equilibrium between the atoms in the gas phase and those in an adsorbed layer with attractive forces between them. The object of these calculations was to develop an equilibrium theory which would help to give an insight into the phenomena observed by Cockcroft (1928) in connection with the effect of stream density on the condensation of metals on solid surfaces. Similar calculations were carried out in 1936 by Peierls using Bethe's approximate method.

We have already considered briefly the application of Peierls's development of Bethe's method when there are repulsive forces between adsorbed particles. It may be mentioned here that Bethe's method gives accurately the form of the isotherm when the repulsive forces are so strong that adsorption is completely excluded on the sites which are immediate neighbours of an occupied site. This application of Bethe's method has been severely criticized by Tonks (1940), but the author (1940 b) has shown that these criticisms are without foundation and that the alternative method of dealing with the problem proposed by Tonks is invalid, except in the extreme limits of a nearly empty or a nearly complete film. This matter will be more fully discussed in a paper by the author which it is hoped will be published shortly.

In general, attractive forces between adsorbed particles will have appreciable effects only when we are dealing with the adsorption of vapours, and with these the possibility of the formation of more than one adsorbed layer must be taken into account. The theory has been extended by Dube (1938) to include this possibility.

Experimental examples of such systems have been obtained by Orr (1939 b), who has made a very careful study of the adsorption of argon, oxygen and nitrogen on crystals of KCl and CsI, and from his accurately determined isotherms has deduced the variation of heat of adsorption as the surface becomes covered. The results for the heat of adsorption of argon on KCl and CsI shown in figure 16 may be taken as typical. The heat of adsorption falls at first and then rises to a maximum. Using Dube's theory, Orr (1939 a) showed that this maximum occurs very near the point at which the first layer is complete. He therefore took the occurrence of the maximum as marking the point at which the first layer was

complete. The initial fall was assumed to be due to the presence of very small cracks which are filled first and in which the heat of adsorption is abnormally high.

The maximum was much more marked for adsorption on KCl than on CsI. and this was assumed to be due to the fact that on the cleavage plane of CsI (body-centred cubic) all the ions are of the same charge, and the electrostatic field so produced induces dipoles in the adsorbed atoms. The repulsive forces between these induced dipoles counteract to a considerable extent the effect of van der Waals attractive forces between the atoms (Roberts and Orr, 1938). In the cleavage plane of KCl, on the other hand (simple cubic), the ions are of different signs and the atoms are adsorbed above the centre of a lattice cell where the electrostatic field is zero (Lennard-Jones and Dent, 1928). Orr has further

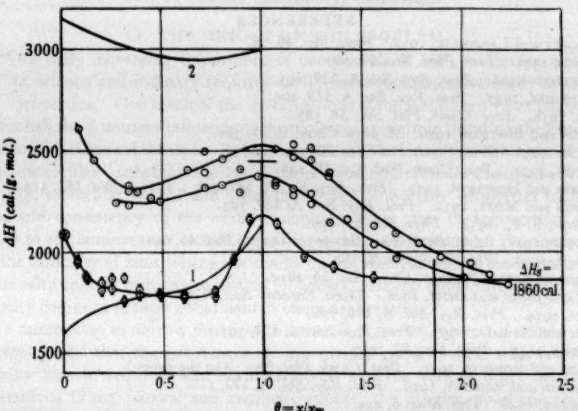

Figure 16. Heat of adsorption as function of θ for argon on KCl (⊙) and on CsI(⊙).

deduced the van der Waals energy of the adsorbed particles from the Kirkwood-Muller formula, using the known atomic constants for the atoms and ions concerned. The theoretical curves for argon, (1) on KCl and (2) on CsI, are shown in figure 16. These curves take into account electrostatic forces as well as van der Waals forces, and it will be noted that, apart from the initial fall, which is due to cracks, they account well for the *variation* of heat of adsorption during the formation of the first layer. For KCl also the absolute magnitude of the heat is given correctly by the theory, and although for CsI the absolute agreement is not so good, it must be regarded as satisfactory.

Orr has further made calculations of the variation of the potential energy of an adsorbed atom over the surface and has shown that in the case of argon on KCl the attractive forces between the adsorbed atoms are so important that

the atoms, which are too large to fit one in each cell, will not follow the square lattice of the solid, but will prefer to form a hexagonal lattice. For details of this and many other interesting points in connection with the field near the surface, reference should be made to the papers by Orr cited, in which references to earlier work are given.

ACKNOWLEDGEMENTS

I wish to thank the Council of the Royal Society and the Council of the Cambridge Philosophical Society for permission to reproduce figures from their *Proceedings*; also the Cambridge University Press, who have permitted me to use figures from one of the *Cambridge Physical Tracts*.

REFERENCES

Blodgett and Langmuir, 1932. *Phys. Rev.* **40**, 78.
Burdon, 1935. *Proc. Phys. Soc.* **47**, 460.
Cockcroft, 1928. *Proc. Roy. Soc.* A, **119**, 293.
Devonshire, 1937. *Proc. Roy. Soc.* A, **158**, 269.
Dube, 1938. *Proc. Camb. Phil. Soc.* **34**, 587.
Eucken and Bertram, 1936. *Z. phys. Chem.* B, **31**, 361.
Fowler, 1936. *Proc. Camb. Phil. Soc.* **32**, 144.
Jackson, 1932. *Proc. Camb. Phil. Soc.* **28**, 136.
Jackson and Howarth, 1933. *Proc. Roy. Soc.* A, **142**, 447; 1935. *Ibid.* **152**, 515.
Jackson and Mott, 1932. *Proc. Roy. Soc.* A, **137**, 703.
Johnson, R. P., 1938. *Phys. Rev.* **54**, 459.
Knudsen, 1911. *Ann. Phys., Lpz.*, **34**, 593; 1915. *Ibid.* **46**, 641.
Landau, 1935. *Phys. Z. Sowjet*, **8**, 489.
Langmuir, 1932. *J. Amer. Chem. Soc.* **54**, 2819.
Lennard-Jones and Dent, 1928. *Trans. Faraday Soc.* **24**, 92.
Mann, 1934. *Proc. Roy. Soc.* A, **146**, 776.
Mann and Newall, 1937. *Proc. Roy. Soc.* A, **158**, 397.
Michels, 1937. *Phys. Rev.* **52**, 1067.
Miller and Roberts, 1941. *Proc. Camb. Phil. Soc.* (In the press.)
Morrison and Roberts, 1939. *Proc. Roy. Soc.* A, **173**, 1.
Oliphant, 1928. *Phil. Mag.* **6**, 422.
Orr, 1939 a. *Proc. Roy. Soc.* A, **173**, 349; 1939 b. *Trans. Faraday Soc.* **35**, 1247.
Palmer, 1937. *Proc. Roy. Soc.* A, **160**, 254; 1938. *Ibid.* **168**, 190.
Palmer and Clark, 1935. *Proc. Roy. Soc.* A, **149**, 360.
Peierls, 1936. *Proc. Camb. Phil. Soc.* **32**, 471.
Raines, Barbara, 1939. *Phys. Rev.* **56**, 691.
Roberts, 1930 a. *Proc. Roy. Soc.* A, **129**, 152; 1930 b. *Ibid.* **129**, 146; 1932 a. *Ibid.* **135**, 192; 1932 b. *Trans. Faraday Soc.* **28**, 395; 1933. *Proc. Roy. Soc.* A, **142**, 518; 1935 a. *Ibid.* **152**, 445, footnote on pp. 447 and 473; 1935 b. *Nature, Lond.*, **135**, 1037; 1938. *Proc. Camb. Phil. Soc.* **34**, 399; 1939. *Some Problems in Adsorption* (Cambridge University Press), pp. 8, 31, 38, 81; 1940 a. *Proc. Camb. Phil. Soc.* **36**, 59; 1940 b. *Nature, Lond.*, **146**, 372.
Roberts and Orr, 1938. *Trans. Faraday Soc.* **34**, 1346.
Soddy and Berry, 1911. *Proc. Roy. Soc.* A, **84**, 576.
Tonks, 1940. *J. Chem. Phys.* **8**, 477.

VISCOSITY AND THE NATURE OF SUBSTANCES OF HIGH MOLECULAR WEIGHT IN SOLUTION*

By F. EIRICH,
Colloid Science Laboratory, Cambridge

§1. THE ORIGIN OF THE PROBLEM

The daily increasing importance of compounds of high molecular weight in science and industry requires an intensive and extensive study of their properties. One result of the investigations of the last fifteen years has been to establish the dominant influence of the molecular weight in deciding the general behaviour of these substances. Recent interest in the viscosity of solutions of the "high-molecular" substances is due to the relation undoubtedly existing between viscosity, molecular weight and other hardly less important properties such as shape and consistency of the units in solution, their state of aggregation, etc. Owing to the comparative simplicity of viscosity measurements on the one hand and the difficulty of interpreting the results on the other, as well as the importance of viscosity control in many industrial processes, one can easily understand how viscosity comes to occupy a vast field in chemical literature.

In attempting to narrow this field to matters of general interest it should first be emphasized that the substances we are mainly going to deal with (such as proteins, starch, cellulose and its derivatives, and the various other materials for artificial fibres, natural and synthetic rubber, the plastics, synthetic resins, synthetic glasses, etc.) form a special group within the general field of colloid chemistry. Colloid chemistry deals in general with bodies or solutions which are in an intermediate state of aggregation between a molecular (or ionic) dispersion and that of microscopic particles. This aggregation corresponds to a turning point in the curve representing the decrease of free energy during a solution or a precipitation process, thus providing some sort of state of a character between an equilibrium of low stability and a very slow rate of change. In other terms, the particles of the body going into solution or separating out of it can at certain dimensions be prevented from further reactions with each other or with the solvent by becoming coated with an adsorbed layer of less active molecules, so that the degree of stability is conditioned by the nature and the stability of this covering.

The molecules involved in such a process are in many cases of an ordinary

* Note by Editor. Dr. Eirich was engaged on this article when he was interned, and it has therefore been necessary to print it without the bibliographicial references.

size, yet recently the fact has been established, by application of various methods of pure chemistry, colloid chemistry and applied physics, that in the case of the above-mentioned group of substances, the so-called macromolecules, the independent units (in the solid as well as in solution) do not originate from a more or less orderly aggregation of smaller molecules, but from giant molecules in which a great number of smaller molecules are linked together by forces of the same nature as those responsible for their own structure. The majority of these compounds, the *high polymers*, are composed of a great number (up to ten thousand) of equal molecules which have been connected by special reactions called polymerization or poly-condensation.

The meaning of this fact was for some time obscured by the circumstance that in solution the molecular weights of these bodies were as far from uniform as are the particle sizes in the colloidal aggregates. Recently, however, a great number of mono-disperse proteins have been discovered by Svedberg, whilst, on the other hand, the imperfection of our methods of preparation has been recognized as responsible in many cases for the existence of these mixtures of homologous polymers. The preparation and purification of the high-polymer natural products is hardly possible without a partial degradation of the genuine molecules, in which the molecules are not equally, but only statistically, affected. Again, the theory of the kinetics of polymerization reactions has proved that the various methods so far in use can only produce molecules of very different sizes. In that connexion the determination of the molecular weight or its distribution has gained the importance of a decisive criterion, as it enables us from the degree of degradation to draw conclusions as to the changes which have taken place, as well as to judge the success of the polymerization process.

It must be remembered, however, that apart from this distribution, the size of the independent units in solution can vary according to the different states of aggregation, because giant molecules have an even greater tendency to build intermediate states between molecular solutions and precipitates than have normal molecules. All that was stated above for the association colloids is therefore also true for the high moleculars, with the sole modification that the single molecules themselves have colloidal dimensions and that the aggregates, if present, may be built up of units of varying sizes.

Attempts to use viscosity methods for the determination of molecular weights originated from the observation by some of the pioneers of a relation existing between molecular weight and viscosity. This observation was adopted by H. Staudinger for checking the magnitude of his polymerization products when he started his well known investigations to elucidate the structure of the high polymers by means of synthesis. From this vast mass of experimental material, Staudinger was able to summarize the relation between viscosity and molecular weight into the formula $\eta_{sp}/c = KM$, where $\eta_{sp} = \eta_{rel} - 1$, M = molecular weight, c = concentration; and K is an empirical constant which has a characteristic value for each series of homologous polymers. Considering the difficulties of the other

method for the determination of molecular weights, we can appreciate the importance of such a simple relation. Actually it renders very good service as a first approximation. It is almost to be regretted that further investigation was unable to prove this formula to be general, but that its validity was shown from critical considerations to be limited. In a later paragraph it will be shown how far these criticisms have contributed to the understanding of the properties of the high moleculars in solution in general and of their viscosity in particular.

§2. THE NATURE OF THE PROBLEM

How far does a direct or indirect relation between viscosity and molecular weight exist? In answering this question one has first to remember the fact indicated by the experimental material that such an alleged relation exists only for solutions of high moleculars with, presumably, asymetrically shaped particles. Also the familiar observation that viscosity is raised by concentration, in as much as the molecular weight is increased, is only valid for solutions of non-spherical particles.

To start systematically, we should deal first with the problem of the viscosity of liquids in general. In spite of remarkable progress during the last ten years, however, we are still far from a satisfactory quantitative understanding of this. Nevertheless, a few preliminary remarks may be useful when reading the following paragraphs. The larger kinetic energy of a liquid compared with that of a crystal increases the distances of the molecules (or single units in general) and facilitates a much greater number of configurations, or a decrease in order. Each time that the thermodynamic fluctuations in the neighbourhood of a particle are extensive enough, there is a possibility of this particle slipping through this opening into another configuration; this opportunity will be used by particles which have sufficient energy to leave the momentary position of equilibrium. This mechanism indicates a close connexion between viscosity and the reaction kinetics.

On this view, the viscosity depends on the velocity of these changes of the local micro-structures and, therefore, on what is called self-diffusion. In a stagnant liquid these changes occur irregularly in all directions. In a flowing liquid those changes of configuration will be favoured which are able to relax the stresses occurring. From this description one recognizes first a series of important properties which have to be taken into account: the cohesional forces (activation energy), the type of co-ordination (shape of the particles and specific volume) and the particle mass (propagation of impulses). In consequence, liquids obviously have to be divided into different classes.

Starting from this general description it is possible to derive quantitative formulae for the viscosity in different ways. Van der Waals, jun., and later Andrade, started from a conception similar to that used in the theory of gases, considering thus especially the Brownian movement and the transference of impulses by the collision of molecules coming from layers with different relative velocities. Prandtl regards the viscosity as a relaxation problem, and assumes

the molecules to be in a static equilibrium which is disturbed by an external stress and makes the particle jump into new positions of equilibrium. According to Bernal and Ward, the degree of disorder which exists in any liquid as a result of the contest between the cohesional forces and the thermal energy is increased by the shearing stress, which supplies the necessary additional energy. The viscosity must be inversely proportional to the number of molecules which, according to a probability distribution, have obtained enough energy to leave the existing configuration and so to relieve the stress. A closely related idea is adopted by Eyring, who assumes that the activation energy necessary for the molecules to jump is to be expressed in terms of the evaporation energy, the effective fraction of which is given by the prevalent type of configuration. Finally, we should like to mention McLeod's method, which starts from the known dependence of the viscosity on the intermolecular volume and extends this relation by the use of the van der Waals equation. Similar considerations were expressed in a more detailed way by van der Wijk and Seeder, who explained the free volume on the basis of the degree of order of the type of configuration. This variety of possibilities for attacking the problem shows impressively how complex the phenomenon of viscosity is even in simple cases.

In conformity with the statistical nature of the problem, all these considerations, even if starting from different mechanisms, have to employ a probability function, and all the various results can therefore be written in the common form, $\eta = Ae^{B/RT}$. Existing differences of opinion refer to the actual value of B and to the interpretation of the factor A according to the assumed derivation. It can certainly be stated that the molecular weight will enter the formula in connexion with the kinetic energy (Entropic factor). Correspondingly, general experience shows that the majority of normal liquids agree with the equation $\eta/M = \text{const}$. For homologous series, in which the molecular shape changes with the molecular weight, another relation holds, namely, $\eta/M^2 = \text{const}$. In this case, however, viscosity can also be expressed as a function of the density, being only slightly affected by the molecular weight. An attempt to take the influence of the shape into account was made by Herzog and Kudar. They found that the increase of viscosity due to the rotation of elongated particles is considerable—in fact, of the same order of magnitude as the transference of impulses by means of the translational motion—and is to be attributed to the fact that the number of particles participating in a special configuration is influenced by steric factors.

For two other attempts to work out the influence of the molecular shape on the mobility and the structure of the liquid, the papers by Eirich and Simha and by Kyropoulos should be consulted.

Turning from pure liquids to solutions, there is a still greater scarcity of quantitative relations, since our general knowledge of their structure is far less developed. The whole situation for solutions is, indeed, so much more complicated because we know very little about the structural changes in a liquid caused by the presence of foreign molecules. These changes must depend on their shape as well as on the

molecular forces, and such changes must be expected even for very low concentrations. Nevertheless it is rather surprising that we know of no binary system in which the viscosity can be expressed as a linear function of the concentration. Ideal mixtures (in the sense of Raoult's law*) provide no exceptions. The viscosity remains in these cases below the straight line, apparently as an effect of the necessarily increased disorder.

In the case of solutions of high moleculars we cannot even expect them to behave ideally in relation to Raoult's law because their relative sizes, as well as their molecular forces, differ vastly from the assumptions under which this law is supposed to hold. Considering the high melting point and heat of fusion of the macromolecules, one would not indeed expect any solubility at all, remembering that dissolved substances in a liquid assume the liquid state, so that the heat of solution is actually a heat of fusion. Finally, whereas a solution is usually defined as a homogeneous phase, solutions of high molecular compounds are necessarily far less homogeneous.

As a matter of fact, the behaviour which is usually exhibited is very peculiar. The two special features are the process of swelling and the lack of a definite solubility. The swelling is a large-scale dissolution of the liquid (or absorption of an equivalent amount of vapour) in the high molecular solid, yielding a spongy body. The parallel decrease in free energy, determined usually by the swelling isotherm, is frequently equal to the heat developed, thus indicating strong molecular interaction. If this heat of swelling, augmented by the possible gain in entropy, is greater than the sum of the cohesional forces of the solvent and high molecular solid, then swelling is followed by a solution of the jelly. As an alternative to this, if the solid is built up of microcrystalline units, a dispersion of these *micelles* can take place, since the cohesion between the units is weaker than between the single molecules of which it consists. Swelling can consequently be considered as a topo-chemical solvation reaction forming a complex molecule which is now miscible with more of the solvent. In other words, as described in detail by Kendall, we are faced with an extreme case of increased solubility by complex formation, in which the solubility of the pure substance is practically zero.

Recently we owe a further elucidation of this problem to Brønsted. In consequence of the fact that the translatory kinetic energy of a very large molecule is negligible compared with the difference in the potential energy in different phases, the distribution of the solute is very extreme. Brønsted points out that in the case where the partition ratio for two solvents for a molecular weight of 100 differs by only 1 per cent, the partition ratio for $M = 500{,}000$ would be 10^{22}. The same consideration holds for the solubility, and it is quite obvious from this tremendous solvent effect that a liquid will be either a very good solvent or strictly a non-solvent. Therefore a solubility in the ordinary sense cannot be determined, and the apparent solubility equilibria occurring must be attributed to a partial dissolution of a substance which is constituted polydispersely or as micelles.

* Van t'Hoff's law is less well-founded.

As a further consequence, a small change of temperature must show the same enormous influence as a small change in the nature of the solvent by impurities or by altering the proportions of a mixed liquid.

Referring to the nature of the solvation, some evidence for the existence of well defined complexes in the earlier stages of swelling was obtained by X-ray methods and given in papers by Katz and by Hess. Recently Hess has interpreted his interesting X-ray work on colloidal solutions flowing through a capillary as indicating an oriented layer of solvent molecules round the particles. Whether such a high degree of order generally exists in solutions and how thick such an absorbed layer might be is not yet known with certainty. Glasstone at any rate pointed out that for solutions of low molecular weight, complexes of solute and solvent must not be regarded as exactly the same in solution as in the crystalline state. For our considerations, however, it is safe to assume that the high-molecular particles are surrounded by a layer of solvent a few molecules thick. This layer, quite different in its geometric and energetic structure, will produce some changes in the surrounding solvent which will decay at not too great a distance.

These conclusions lead to an interesting consequence. As can be shown by elementary geometry, the distances between the dissolved particles at a given concentration are independent of the particle size if the distances are expressed by means of characteristic dimensions of the solute. If, for instance, R is the diameter of equal spherical particles, then the mean distance between them in a 1 per cent solution is approximately $3R$. If R is of the order of magnitude of the molecules of the solvent, then only a few of them come to lie between the solute particles, and the influence of the latter will be very great. In colloidal solutions R can be a thousand times larger, and consequently larger amounts of solvent can remain practically unchanged. We are here faced by the necessity, which is characteristic of our problem, of defining the conception of concentration much more sharply so far as mutual interactions are involved. Furthermore, it is because of the dimensions of the particles that the character of the solution becomes more and more similar to a two-phase system, which is in good agreement with the observed thermodynamic peculiarities.

The large dimensions also account for the existence of the strange mechanical behaviour commonly found, if they are combined with marked deviations from a regular shape. Great distances can be then spanned by the asymmetric particles, thus giving rise to a superstructure which is responsible for a great many of the anomalies of flow, the phenomenon of gelation and others.

From the fact that the colloidal systems can be regarded as two-phase systems, it is possible to derive a new and special treatment of the viscosity problem, namely, the hydrodynamic treatment.

§ 3. HYDRODYNAMIC TREATMENT

As this method has proved so far to be the most effective, it will be explained in somewhat more detail. The hydrodynamic procedure necessitates a substantial idealization of the real conditions. In contrast to the molecular conception

of a liquid from which we started, the solvent is now regarded as a continuum, and Euler's equations instead of those of Lagrange are used for describing its movements. Furthermore, it is always assumed that the dimensions and movements of the particles are small enough to justify the neglect of mass forces, that the particles are rigid and of rotatory symmetry and that the liquid is infinite and adheres to the surface of the particles. As attention is confined to liquids of Newtonian character, the velocity components should satisfy the Stokes-Navier differential equations, of which, for small Reynolds numbers, the terms of inertia can be neglected.

From these equations one obtains as the amount of kinetic energy transformed into heat by friction in the unit of time

$$\frac{dW}{dt} = 2\eta \iiint a_{ik} a_{ik} \, d\tau, \qquad \ldots\ldots(1)$$

where a_{ik} are the velocities of deformation,

$$\frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right),$$

and the integral has to be taken over the whole volume of the liquid. Equation (1) defines η, and is used for the calculation of the viscosity of the suspensions.

The first to investigate the increase of viscosity of liquids by suspended rigid spheres was Einstein. In that case the transport of impulses covers a distance equal to the diameter of the spheres instead of a distance of the order of magnitude of the free path in the solvent, resulting in an increase of viscosity. In connexion with this, the following points are further relevant. The motion of a liquid in which the velocity components have a linear dependence on the Cartesian coordinates can be interpreted as consisting of a translation, a rotation and a dilatation. For any other type of flow this is only true for the relative velocities of two infinitesimally neighbouring volume elements. Linear motion, for instance, is realized between two parallel moving plates whose relative distance stays constant. Only in this case can Einstein's results be applied. From the postulated absence of inertia terms it follows, for reasons of symmetry, that only the dilatation part of the undisturbed flow has to be taken into account in solving the problems for the limiting values of spheres, since rotation and translation are simply followed in this case. Taking the viscosity η of a suspension of volume-concentration c in a solvent of viscosity η_0, Einstein thus obtained the well known relation

$$\eta = \eta_0(1 + 2{\cdot}5c). \qquad \ldots\ldots(2)$$

This means that the increase in internal friction is independent of the absolute size of the suspended spheres. This is intelligible in view of the neglect of the effects of concentration and inertia.

Jeffery treats the same problem under the same assumptions for ellipsoids. A substantial complication enters, however, owing to the rotation of the particles relative to the liquid. According to the former assumptions, accelerating forces and turning moments must not act on the particles. By fulfilling these postulates

one obtains the equations for the spin of the particles. Introducing Euler's angles θ, ϕ, ψ for ellipsoids of rotation with semi-axes a_1 and a_2, they have been formulated (H = rate of shear):

$$\left.\begin{aligned}(a_1^2+a_1^2)\dot{\phi} &= H(a_1^2\cos^2\phi + a_2^2\sin^2\phi),\\ (a_1^2+a_2^2)\dot{\theta} &= \tfrac{1}{4}(a_1^2-a_2^2)\sin 2\theta \sin 2\phi,\\ \dot{\phi}\cos\theta + \dot{\psi} &= \frac{H}{2}\cos\theta,\end{aligned}\right\} \quad \ldots\ldots(3)$$

with the integrals

$$\tan\phi = \frac{a_1}{a_2}\tan\left(H\frac{a_1a_2}{a_1^2+a_2^2}t\right); \quad \tan^2\theta = \frac{a_1^2a_2^2}{\beta^2(a_1^2\cos^2\phi + a_2^2\sin^2\phi)}.$$

The constant which occurs in the integration of the first differential equation is determined by taking $\phi = 0$ for $t = 0$. At any moment the angle θ appears to be determined by ϕ, and owing to this relation the second constant k cannot be eliminated. The spinning axis describes an elliptical cone, the aperture of which varies between $\tan^{-1} a_1/k$ and $\tan^{-1} a_2/k$. Any value of k would be in accordance with the fundamental hydrodynamical equations, and therefore it cannot be determined from these equations in their present form. This indeterminacy enters into the expression for the energy dissipated, and hence into that for the viscosity of the suspension. In order to avoid these difficulties, Jeffery makes the additional hypothesis that an orientation will occur, at which the dissipated energy is a minimum. In that case elongated particles should orient themselves with their axes at right angles to the plane of flow ($\theta = 0$), whereas a position in the plane of flow, with ϕ as a periodic function of the time, would correspond to a maximum of dissipated energy. The converse holds for oblate ellipsoids of rotation. In the limiting case of $a_1/a_2 \gg 1$, one obtains

$$\eta_{\min.} = \eta_0(1+2c); \quad \eta_{\max.} = \eta_0\left[1+\left(\frac{f}{2(\log 2f - 3/2)}+2\right)c\right]. \quad \ldots\ldots(4)$$

As will be shown by experiments described below, short rods behave in accordance with Jeffery's assumption. Particles possessing high axial ratios turn over into the plane of flow, where they orient themselves primarily parallel to the velocity vector. In this position the ϕ component of the angular velocity reaches a minimum which becomes more pronounced with increasing elongation. It should be mentioned that in this position the energy dissipated has another minimum. Up till now, however, the theory has not been able to give an unambiguous solution of the problem of the orientation. It has been shown in greater detail elsewhere that the introduction of the forces of inertia cannot be expected with certainty to eliminate the indeterminacy. It seems therefore necessary to investigate the problem of the non-stationary case in order to see how a certain orientation can be reached. It seems very difficult to achieve this task, which Eisenschitz and Burgers tried to avoid by assuming an isotropic initial state; but as "initial state" here means the beginning of the stationary case, it is difficult to understand an isotropy at this instant. On the other hand, the stability of the two limiting positions

found in the experiment has been studied by Simha. By employing the method of small oscillations it became intelligible why the positions perpendicular to the plane of flow lose their stability with increasing values of $f(Ua_1/a_2)$.

The Brownian rotary motion counteracts the orientation of the flow when the particles become sufficiently small, and the distribution function of the positions has to be calculated. Problems of this nature were first treated by Einstein. Langevin and Debye treated an analogous case in studying the question of the orientation of magnetic or electrical dipoles in an external field, where the solution of the corresponding Plank-Fokker equation is given by a Maxwell-Boltzmann distribution. It is not likely, however, that this applies to our case, where non-conservative forces are acting. In any case it has to be borne in mind that the problem, as long as hydrodynamic forces are effective, reduces to one plane, since the variables θ and ϕ are then not independent of each other.

The distribution function mentioned above can be based on a type of generalized diffusion equation. In developing the distribution function further, the diffusion coefficient D is expressed as a function of the axial ratio. A quotient H/D (H = rate of shear) is obtained in the explicit formula, the magnitude of which characterizes either the prevailing Brownian movement ($D \gg H$) or the hydrodynamic orientation ($D \to 0$). Difficulties arise again if one has to deduce the average value of the dissipated energy and to consider the distribution function in each case. As a consequence $d\overline{W}$ remains ambiguous, so long as the coupling of θ and ϕ prevents elimination of the integration constant. This is only the case for overwhelming Brownian movement. Reports on this difficult problem have been published quite recently, and we shall revert to it in a later section. Here it need only be mentioned that Burgers (following a similar method to Eisenschitz) obtained the equation

$$\eta_{sp} = \eta_{rel} - 1 = \frac{4\pi a_1 a_2^2 V}{3}\left[\frac{f^2}{15} - \frac{1}{\log 2f - 1{\cdot}80}\right]c, \qquad \ldots\ldots(5)$$

while Kuhn derived in an elementary manner the approximate formula

$$\eta_{sp} = \left(2{\cdot}5 + \frac{1}{16}f^2\right)c. \qquad \ldots\ldots(5\,a)$$

For a partial Brownian movement, Burgers, following on the work of Boeder, obtained an approximation formula in which the viscosity was proportional to a_1/a_2 for the suspended ellipsoids.

In conclusion it should be remarked that according to the hydrodynamic treatment the rise in viscosity caused by suspended particles can be generally expressed by a formula of the type $\eta_{sp} = \gamma . c$, where γ assumes values depending on the particle shape and the intensity of the Brownian movement, of which the above-mentioned formulae (2), (4) and (5) gave some examples. From this, comparing the viscosities of a homologous series of thread-like molecules, conclusions can be drawn as to the molecular weight. In such a case the volume,

length and molecular weight of the particles can be assumed as proportional and the equations above be transformed to

$$\eta_{sp} = K_1 M^2 c \quad \text{or} \quad \eta_{sp} = K_2 M c, \qquad \ldots\ldots(6)$$

according as the Brownian motion is dominant or negligible. (For the intermediate case, see the discussion below.) It is obvious, although rather surprising, that Staudinger's rule coincides with the second formula.

There still remains to be pointed out the part which is played by the solvation jacket round the particles. This coating not only renders them soluble and justifies the hydrodynamic treatment, but has to be added to the particle volume or to the concentration. Before any comparison between a calculated and a measured viscosity can be made, the amount of bound solvent must be known. On the other hand, if a formula can be assumed to hold, this amount can be determined. It would be beyond the limits of this report to enter into discussions of this important field. It need only be emphasized that the results derived from viscosity measurements under certain conditions agree with those of other methods, and yield values for the amount of bound solvent of the order of 50 to 100 per cent of the dissolved dry weight.

§ 4. PAPERS ON HYDRODYNAMICAL ASPECTS

In this section we shall discuss some papers briefly without claiming to make an exhaustive survey. We start by mentioning a series of papers by Peterlin who, starting independently from similar considerations to those of Eisenschitz and Burgers, obtained results essentially confirming those of his predecessors. He takes the rotary diffusion constant as given by Gans as a measure of the intensity of the Brownian movement, and determines the diffusion function in the usual way, calculating the difference between the numbers of particles which migrate into an elementary volume per second with and against the current. For this purpose he uses Jeffery's equations. The resulting general diffusion equation can be divided into two parts, one independent of and the other dependent on the time. The latter serves for the description of the periodic fluctuations of the distribution during the start of the flow. The other part he uses for the determination of the partition in the stationary case, and according as diffusion or the rate of shear is the greater he obtains his final distribution function. Peterlin draws further some interesting conclusions for the intermediate range between extreme orientation and isotropic distribution. As the viscosity depends here on the axial ratio as well as on the quotient H/D, a dependence of the viscosity on the rate of shear (occurring in many experiments) can be calculated and is found to be more pronounced when a_1/a_2 or the absolute dimensions of the particles grow. There is a limitation of this region given by the same dimensions, in so far as for small particles no rate of shear can be realized which is sufficient to orient them, while for larger particles the time necessary to reach their final positions becomes so long that as a result the viscosity becomes indeterminate.

Criticisms can be urged against this paper in the light of the preceding sections.

In the hydrodynamic case (vanishing Brownian movement) the distribution function is doubtless given by $1/\phi$, which also follows from the plane diffusion equation for $D=0$. Peterlin, however, obtains his results by means of a spatial equation. Considering his treatment of Brownian movement, the quantity which had to be averaged in Debye's case was independent of the type of field and affected only by its direction. Jeffery's equations, however, as well as his value for the energy dissipated, were derived under the assumption of vanishing forces, which involves a coupling of θ and ϕ. The distribution function has only to deal with ϕ, and k* still enters in dW/dT so long as no further assumption about the orientation is made. In the case of complete Brownian motion, no regular rotation in the plane of flow occurs, and in this case θ and ϕ have not to be coupled. As it is not correct to assume that the particles move in this case with a hydrodynamic velocity, Peterlin is no longer justified in using Jeffery's expression for the dissipated energy

The transition from an indeterminate k to its elimination by diffusion is likely to take place continuously. So long as the diffusion does not exceed any other influence, the particles will change their k values frequently, but in the intervals will continue their hydrodynamic movements, as all positions are equally possible. As the diffusion becomes more predominant these fractions of the hydrodynamical path tend to become negligible. An attempt to limit these cases was undertaken by Burgers, whereby, however, the coupling of the two angles was again left out of account. He took as criterion the degree of convergence of a series determining the frequencies of orientation in terms of H/D.

Finally, some complications of Peterlin's solution of the plane-diffusion equation should be mentioned. The solution cannot be a Maxwellian one because the potential is not conservative. It contains two constants, of which one was determinate in Debye's case. In making his approximations in calculating the effective momentum, Peterlin obviously tries to avoid this difficulty.

Before we start to discuss an interesting paper by Huggins we must mention briefly the work of Kuhn on the statistical shape of chain molecules. Assuming free rotation around the C–C bond in a polyatomic chain of C atoms, statistical considerations show that the chain will have a zigzag form similar to the path of a molecule in a Brownian movement. The mean length can be calculated and is found to be equal to $AK\sqrt{N}$, where A is a numerical factor, N the number of C atoms and K another factor which is determined by various other data, for instance valency angle, ratio of the energy necessary to effect rotation to the energy which is actually available (RT), etc. This important result is, of course, only a first approximation, as the volume of the atoms themselves which compose the chain is neglected, together with their mutual attraction. Nevertheless, it will seriously affect the relation just derived between molecular weight and viscosity of thread molecules. The coiling up will also exert a second influence, namely, on the volume of the dispersed phase, because the volume of the coils will increase

* The integration constant mentioned before, which governs the opening angle of the cone along which the ellipsoids rotate.

more quickly with the chain length by the inclusion of liquid, in contrast to stretched molecules, where the volume grows proportionally to the number of atoms.

Huggins extended Kuhn's work to bead chains of unspecified length as well as to bent molecules. In the first case the internal friction caused by the relative movement of the spheres of a dumb-bell in a shearing liquid is calculated again, but now summed up for i components. Consequently the theory of Boeder and Kuhn dealing with the distribution of the particle axes in a hydrodynamic field disturbed by Brownian movement is used and extended to the polyatomic molecule. He finds that in the case of rigid rods $\eta_{sp} = \frac{N}{12} \cdot \frac{a_1}{a_2} \cdot c$ (N = number of molecules in the chain), a modification of Kuhn's result. Introducing the statistical average form with the assumption that molecules remain rigidly in this state, he finds, however, a direct proportionality of the viscosity to the number of the atoms in the chain, that is, he establishes Staudinger's law. In conclusion Huggins makes some interesting remarks about the influence of a possible flexibility. In this case the strains and stresses exerted alternately on the particle by the surrounding liquid will be partly compensated by the consumption or the gain of energy which is effected during the changes of the existing configurations and, therefore, of the entropy. The flexibility factor does not deviate much from, and is always less than, unity; it increases with the temperature and decreases with the length of the particle or the rate of shear.

From the point of view of the theory one must object to the use of Kuhn's model. Another drawback of Huggins's treatment is that he follows previous authors in assuming that the regular hydrodynamic motion is maintained even under the influence of strong Brownian movement. Nevertheless, his interesting attempt to formulate the influence of flexibility is very welcome, although he does not at present give any details of his calculations. It is thus impossible to judge the full importance of his result. He considers apparently only the change of shape, but does not take into account the fact that a particle, being free to perform internal movements, may disturb the flow of the neighbouring liquid less than a rigid particle.

Another attempt to solve the problem of the Brownian movement was made by R. Simha. He regards the case of overwhelming Brownian movement only, in which case, in his opinion, the macroscopic flow is no longer able to bring about a regular rotation of ellipsoids, but gives way to a uniform distribution of the axes. For this reason, and here is the decisive step, a change must be made in the method of expressing the dissipated energy as a function of the orientation. The formula hitherto used was derived with the aid of Jeffery's equations of motion (3), and involved implicitly the coupling of the two angles θ and ϕ. Now, without a mutual dependence of these angles and without a regular rotation, the former considerations have to be modified to the effect that the distribution function $f(\phi) = N/4\pi$, and that the increase of dissipated energy is due to ellipsoids isotropically

oriented, which no longer follow the motion of the solvent. Jeffery's procedure can then be repeated with these modifications. Simha obtained the formula for suspensions in a rotating cylinder apparatus and for axial ratios much greater than unity:

$$\eta_{sp}/c = \frac{f^2}{15(\log 2f - \frac{3}{2})} + \frac{f^2}{5(\log 2f - \frac{1}{2})} + \frac{14}{15}. \qquad \ldots\ldots(7)$$

The first term is the same as that derived by Eisenschitz and Burgers, while the additional increase is due to the relative movement of the particles and liquid. The approximately quadratic dependence of η_{sp} on the axial ratio for rigid rods is preserved. It should also be noted that Simha's result gives the steepest increase of all formulae. As regards the orientation (even when small) indicated by the streaming birefringence, the question of how far his assumptions are justified remains.

We come now to some experimental papers related to the hydrodynamic theory. First, the continuation of the series of model experiments which were started by the author and his co-workers should be mentioned. As is evident from the discussion above, the theoretical results cannot easily be applied to the complex systems which occur in practice, since up to date there hardly exists an independent method by which it could be ascertained to what extent the given properties correspond to the theoretical assumptions. Therefore the only possibility of testing or amending the theory appears to be by the use of model suspensions which permit a satisfactory definition of all the required magnitudes, as well as their arbitrary variation. Obviously, such systems cannot and should not represent all properties of natural systems.

The first papers dealing with methods, preparation and the testing of Einstein's and Jeffery's equations have been repeatedly discussed. The new paper by the author and Sverak repeated and completed parts of the previous work and undertook extensions for Brownian movement and flexible particles.

Amongst the results with rigid rods, the orientation should be primarily discussed, since it touches upon the crucial problem of the theory. It should be remembered that in the Couette and similarly in the Poiseuille (capillary) flow, two prevailing orientations were found, one in the plane of flow ($\theta = 0$) parallel to the velocity vector and the other at right angles to this plane. According to the theory and earlier experiments of G. I. Taylor, the particles are in a state of rotation and precession. The new cinematographic record of the orientation of the particles in the Couette apparatus shows quantitatively that with increasing axial ratios the particles preferably lie horizontally. The uncertainty of the theory does not appear in the experiment, in contrast also to Peterlin's derivation, and no dependence on the time or on the rate of shear was observable in these experiments. In the capillaries the viscosity of longer rods was found to vary in the way postulated by Peterlin for smaller particles. (According to Simha, the results of the theory should hold for capillaries if concentration and velocity are low enough.)

The author and Sverak also found values for the specific viscosity in better agreement with Jeffery's formula than in the former paper. This was partly due to an improvement in the methods, but is mainly to be ascribed to the improved possibility of extrapolation, supplied by the formula lately derived by Guth and Gold (see eq. (10), p. 350) which is an extension of (4) and gives an approximation for a moderate degree of concentration. By this method it could be further shown that the range of linear increase of viscosity with concentration shrinks rapidly with increasing axial ratio. For $a_1/a_2 = 20$, for instance, the limit is found at a specific viscosity as low as 0·025 and quickly ceases to be measurable at greater elongations of the particles. Taking the quadratic term of formula (10) into account, the limit of validity is again reached very soon, e.g. for $a_1/a_2 = 140$ at $c = 0{\cdot}015$ per cent, that is, when $\eta_{sp} = 0{\cdot}004$. Hence the difficulties here in proving the theory lie in the presence of strong mutual interaction between the particles even at the lowest measurable viscosities.

In model suspensions with Brownian movement (crystal needles of dimethyl glyoxime) only a few measurements could be carried out which satisfied all the requirements, and the thermal motion of the particles was still very slight compared with the hydrodynamical forces of orientation. Nevertheless, the viscosities were considerably higher than those of analogous suspensions without Brownian motion. The increase in viscosity with concentration is greater than in the case of longer rods. In dilute solutions the influence of the apparatus employed or of the rate of shear is small.

It may be regarded as the principal result of the experiments with rigid models that these show the peculiarities of the solutions of high polymers in an even more marked manner. This indicates that really rigid rods are but rarely realized in nature and that in general the particles are "softer". For this reason the experiments with flexible particles (silk threads softened by a xanthogenate reaction) are of special interest. It was found that the viscosities of dilute suspensions of threads are considerably lower than those made from rigid rods. In particular the effect of concentration on the viscosity is markedly reduced. Both in dilute and concentrated suspensions the effects are independent of the apparatus and become more pronounced at higher degrees of swelling, i.e. at increasing softness of the threads. In fact, the viscosity of a thread suspension can drop to that of a suspension of equally concentrated rigid spheres and may even give lower values at higher concentrations.

For the interpretation of these results it was assumed that the calculated axial ratios a_1/a_2 had been reduced by the coiling. The orientation of the particles relative to the direction of flow will also have been changed—obviously the same factors as were considered by Huggins. The reduction of the effect of concentration, however, can be explained on the basis of Taylor's calculations. If soft particles are used, the velocity of the liquid adhering to them need not be zero, as their surface can yield, and the flow of the liquid can produce deformations in their interior. The impulses will not be transmitted over the whole of the

diameter, as for rigid particles, but will be partly absorbed inside. Taylor calculated factors for the relative viscosities of suspensions of liquid spheres and found a decrease from 2·5 to 1 according to the viscosity of the suspended liquids. Similar reductions will have to be taken into account for flexible rods. If once the linear term for the{concentration, viscosity}relation be deduced, the same value will obtain for the terms governing the interaction.

It may be doubted how far the actual conditions are reproduced by these model suspensions, but the significance of the results emerges from the fact that flexibility and Brownian motion counteract each other in a similar way to that forecast by the theory and indicated by Staudinger's results. Moreover, attention has been drawn to the mechanism which probably plays an important part in the complicated relations between viscosity and concentration.

While on the whole these results give satisfactory proof of the justification of the hydrodynamical considerations, the peculiarities in the capillary viscosimeters gave a starting point for the theoretical treatment of the influence of inertia and the type of flow in two further papers by Simha, the author and O. Goldschmidt. Experiments carried out gave some details as to the nature of the disturbances due to inertia, such as a relative translatory and rotatory motion of the suspended spheres and a certain concentration along the axis of the tube, a noticeable influence of the wall appearing to be active in this case.

From the various attempts to apply the derived formulae to experimental results, a paper by Polson is of special interest. Polson plotted the specific viscosities of eleven proteins against the squares of their axial ratios, which he calculated from Perrin's formula for the ratio D_0/D ($=F/F_0$, Svedberg's shape factor) determinate by the combination of diffusion constants and sedimentation constants from the ultra-centrifuge.* He finds a linear relation

$$\eta_{sp}/c = 4 + \frac{1}{10}\left(\frac{a_1}{a_2}\right)^2,$$

which becomes identical with Kuhn's formula (5 *a*) if 1·6 is taken as a factor for solvation. It is of great importance that it should be possible to calculate molecular

* Herzog and Kudar, in the paper already mentioned, and recently Perrin, have derived a formula for the diffusion constant of elongated particles by introducing the average of the resistance experienced by ellipsoids which change their position constantly and at random with respect to their direction of movement through the liquid. If this diffusion constant D is compared with D_0, calculated for particles of the same volume, but spherical shape, the following ratio is found :

$$D_0/D = \sqrt{(1-a_2^2/a_1^2)} \Big/ \left[\left(\frac{a_2}{a_1}\right)^{2/3} \log\left\{1+\sqrt{\left(1-\frac{a_2^2}{a_1^2}\right)}\Big/\frac{a_2}{a_1}\right\}\right].$$

This idea is equivalent to the earlier conception of Svedberg, who derived a shape factor from the measurement of his sedimentation velocities in the ultra-centrifuge by forming the ratio F/F_c, where F is the resistance experienced by the sedimentation particles, determined together with the molecular weight M in combination with diffusion measurements :

$$s = M(1-V\rho)/F; \quad D = \frac{RT}{N}\cdot\frac{1}{\tau} \quad (s = \text{sedimentation velocity}),$$

and F is the calculated Stokes resistance :

$$F_0 = 6\pi\eta(3MV/4\pi N)^{1/3}.$$

weights and axial ratios from viscosity and diffusion data. It must be stated, however, that Polson's formula cannot be used for accurate determinations, because it is still of an empirical character; in the cases under consideration a_1/a_2 can be taken as proportional to D_0/D (or Svedberg's ratio F/F_0), with the result that after eliminating $(6\pi\eta)^2$ a linear relation can be established between η_{sp} and r^2/r_0^2. In other words, the specific viscosity of the proteins is practically a linear function of the quotient of the square of an effective "diffusion radius" to the arbitrary radius $r_0 = (3MV/4\pi\eta)^{1/3}$. Yet various circumstances are against a numerical calculation of an axial ratio or the molecular weights: Perrin's equation assumes rigid, uncharged, rotatory ellipsoids, and it is extremely doubtful whether these three assumptions are fulfilled by proteins; the same is true for Kuhn's approximation, and it seems rather inconsistent to find a uniform solvation of 60 per cent indicated by its adaptation, without considering this fact in the calculation of D_0; the shape of the particles, their charge and their consistency influence diffusion and viscosity together, and they cannot be separated in this manner. Disregarding this fact, and calculating a summary axial ratio, one is bound to obtain too large values, as can be seen clearly by a comparison with reliable data from x-ray studies (Crawfoot, Perutz). Polson, for instance, calculates for haemoglobin an a_1/a_2 of 3·8 instead of 1·5, and for insulin an a_1/a_2 of 3·2 instead of 0·7.

As a matter of fact it should be possible in some cases to obtain better values by this method, particularly if the solvation, flexibility and other influences are negligible compared with the influence of elongation. In that case we should not forget that practically all measurements are carried out in capillary viscosimeters under very high rates of shear, which amongst other difficulties makes it doubtful whether the formula for a complete Brownian movement can still be used. This is true, for instance, in the attempts of Frampton and Neurath and of Lauffer, who tried to calculate the axial ratio of the tobacco mosaic virus in the same way as Polson and found almost twice the value.

In this connexion a paper by Robinson is of great importance. He measured in a Couette apparatus the viscosity of tobacco mosaic virus suspensions and simultaneously the degree of orientation by means of the double refraction of flow. By this procedure the problem of orientation is solved experimentally and the viscosity can be immediately derived as a function of the distribution of the particle axis or, consequently, of the rate of shear. The angle of the isocline ψ and the amount of the birefringence Δ, expressed as a fraction of the maximum possible amount Δ_m calculated by means of Kuhn's formula, are used as a criterion. The viscosity is then calculated, following Kuhn's method, in an elementary way. It is assumed as an approximation that the viscosity should be additively composed of a part originating from the number of particles which remain oriented at random and another part which is oriented in the direction of the isocline, and is at rest relative to the shearing liquid. Robinson's relation can be written as

$$\eta_{sp}/c = \frac{1}{4}\left(\frac{a_1}{a_2}\right)^2\left[1-\frac{\Delta}{\Delta_m}(1-2\cos^2\psi)\right]. \qquad \ldots\ldots(8)$$

A further interesting contribution is the derivation of the temperature dependence of the birefringence, which is negative, in contrast to the positive dependency of the viscosity. It is to be regretted that his numerical results cannot be used to check his calculations, because, using aged preparations, the actual values for the axial ratios of his (aggregated) particles are unknown. In spite of its nature of a rough approximation, Robinson's attempt is very welcome as stimulating, with the assistance of experiments, the further investigation of the problem of interacting Brownian movement and hydrodynamic forces.

§ 5. STAUDINGER'S RULE

When we now try to correlate the results so far discussed with Staudinger's rule $\eta_{sp}/c = KM$ it is obvious that this problem must be divided into different sections. The first question would be how far the rule itself can be considered as established, the second whether it is possible to apply the hydrodynamic methods, and finally, what conclusions can be drawn from such a treatment of Staudinger's material.

Regarding the first point, nothing very important has been added recently to our knowledge. Staudinger's rule seems to be at any rate fairly well established for medium-sized particles. His extrapolation for high molecular weights still remains doubtful, although the work by Schulz on osmotic pressure apparently proves a fair validity also in this range. As Schulz's data are not free from new assumptions about the effect of higher concentrations, for which alone it is possible to measure osmotic pressures when the molecular weights are greater than 100,000, no final decision will be achieved without more independent material, provided by diffusion and sedimentation measurements. The present data partly disprove the rule, while others are in quite good agreement. This inconsistency is largely due to the fact that, even when the same solution was investigated, different averages were bound to be found according to the different way in which the molecular distribution of any fraction influences the osmotic pressure, the rate of diffusion or of sedimentation, the viscosity, etc. As the majority of Staudinger's determinations of molecular weights by viscosity measurements are based on a comparison with osmotic pressures, this relation could be reduced to a form $\eta_{sp}/c \simeq c/p$. This relation obviously depends on a fairly constant molecular distribution in the various fractions, as the osmotic pressure is sensitive to small amounts of highly dispersed material, whereas the average viscosity is affected by a small number of molecules of extreme shape.

If we now consider whether the hydrodynamic procedure is applicable to Staudinger's systems, the influences of solvent and concentration are of primary importance in deciding whether single molecules or aggregates are present, and will be discussed in the last section. If the molecules can be assumed to be single, the next question is whether their size, their shape or the type of flow is suitable for the theoretical treatment. These conditions are to some extent fulfilled by the fact that Staudinger's rule is only applied within one and the same series of high-polymer products, measured under identical conditions. Minor deviations from

the hydrodynamic assumptions about the shape will be covered then by a sort of effective sphere which finds its expression in the specific constant K and remains, as the experiments indicate, constant over a greater range of molecular weights. The use of capillary viscosimeters is of course a serious obstacle to hydrodynamic interpretations, especially for higher molecular weights.

The results of Meyer and van der Wijk are interesting in another connexion. These authors measured the viscosity of the paraffins C_{17} to C_{34} extremely carefully but failed to obtain a linear relation between viscosity and molecular weight. This result is not competent, however, to disprove Staudinger's rule, as Burgers pointed out, because some of the relative viscosities of the paraffins lie even below the values one would expect for spheres according to Einstein's law. These paraffins obviously lower the viscosity of the solvent (carbon tetrachloride) by disturbing its structure and thus violating one of the basic assumptions of the hydrodynamic theory. There are some similar cases known, for all of which it can be stated that the solutions are not ideal and that the molecular weight of the solute is rather small. In view of what was said in the introduction it is this combined influence on the structure of the solvents which tends to limit the use of hydrodynamics, while plenty of cases are known of a fair agreement for smaller (and, of course, for larger) particles.

Staudinger, as is well known, himself tried to interpret his rule in a primitive hydrodynamical way by employing formula (2) under the assumption that the molecules were rigid rods and "occupied" a much larger volume than their own on account of the rotatory Brownian movement. Recently the problem has been carefully examined by Burgers. He uses Staudinger's data for polystyrenes, one of the most important high polymers, with the following constitution (n = number of the basic units): $\left[\begin{matrix}-CH-CH_2-\\ \quad\ C_6H_5\end{matrix}\right]_n$, a length of $2{\cdot}6n$ A. and a mean thickness of 8 to 15 A. To use independent and reliable molecular weights for the comparison, Burgers took values which were determined either cryoscopically or by means of the ultra-centrifuge. The latter values come from a paper by Signer, who measured the double refraction of flow for equivalent samples. These data, together with the fact that the viscosity is independent of the velocity gradient, go to show that formula (5) may be applicable, thus enabling the axial ratio a_1/a_2 to be derived. The known volume of the polystyrenes in solution enabled him finally to obtain a table from which the following data are reproduced:—

η_{sp}/c	M	n	$f=a_1/a_2$	a_2 (A.)	$\frac{\eta_{sp}}{cM}$
7·03	2 460	24	13	8·5	0·00286
13·6	7 600	73	21	10·0	179
12·8	13 000	125	29	11.0	168
25·9	30 000	288	33	14·0	86
54·0	80 000	770	52	16·5	67
230·0	270 000	2600	122	185·5	86

While Staudinger's own data for this case already show many fluctuations, we learn from the above table that there is a rather poor proportionality, if any at all, and different constants are yielded for the middle and high molecular weights calculated by means of (5). The values for the cross-section $a_2^2\pi/4$ appear to increase as $\sqrt{a_1}$, while the length a_1 is almost proportional to $M^{1/2}$. The axial ratio consequently grows still more slowly. These results are in quite good agreement when compared with Signer's values for the double refraction. Data from the ultra-centrifuge, however, indicate an axial ratio proportional to a still lower power of M.

As a result of his examination, Burgers came to the conclusion already arrived at by Kuhn, Mark and collaborators that if a proportionality is found between viscosity and molecular weights, it must be due to a non-linear relation between the latter and the length of the particles, for viscosities are only related to the axial ratios, which in turn lag behind proportionality with the molecular weight. The non-linear mechanism is very probably to be ascribed to a crumpling or a coiling of the elongated molecules, confirming the Kuhn-Huggins theories and in agreement with the model experiments.

At present it is not possible to state anything more, because there are too few measurements covering all the necessary data. There is, however, some more independent evidence against a straight form of the molecules given by other methods. One of these, the measurement of the double refraction, has already been mentioned. The same measurements indicate, furthermore, that the presumed coils undergo a stretching under the shearing force of the surrounding liquid. Other evidence comes from measurements of the diffusion constants, the course of which can be fairly well relied upon to reflect the change of the axial ratios. The results, for instance, of Polson's very thorough investigations of cellulose derivatives show that they can be interpreted by the assumption of an increasing deviation from the straight form with increasing molecular weight. Still less compatible with a straight form are the asymmetry factors obtained by means of the sedimentation constants. Actually they often change so little with the molecular weight that one has to think about additional effects of concentration or flexibility. It was only lately that Signer found one case where the shape factors for methyl cellulose increased as predicted by the theory. But even in this case at least an S-shape of the molecules appears because the increase of length is only about 70 per cent of that of the molecular weight. Moreover, the molecules investigated are not really high molecular (14 000 to 38 000).

The problem of the coiling up of thread-like molecules has recently been dealt with to some extent by Sakurada. Starting from Kuhn's formula, he introduced some rather drastic simplifications and arrived in a less conclusive manner at about the same results as Burgers. Sakurada evaluates his results further by calculating the ratios of the hydrodynamic to the stereo-chemical lengths. These "crumpling factors" are for most substances between 1 and 10, and indicate, therefore, a moderate degree of bending.

§6. EFFECTS OF CONCENTRATION

In a search for the mechanism of the effects of concentration on the viscosity, one meets very divergent facts. Even the dividing line between a dilute and a concentrated solution is difficult to draw. A linear relation, the usual criterion for dilution, is not sufficient in this case because increasing concentration not only starts the interaction but frequently changes the structure of the particles themselves. The papers of Bredee and Booys, Eirich and Mark, Guth, Howink, Philipoff, Sakurada and Ostwald, attempting to make the general problem clear, may be summarized in the following analysis:—

One group of colloidal solutions behaves "normally", that is, a solute of low molecular weight and of the same concentration would raise the viscosity by about the same amount. Many investigations show that the solute particles do not differ much in this case from a spherical shape, nor is their density changed in comparison with the solid. A second group raises the viscosity much more than the same concentration of smaller molecules would do, and the range of the linear relation is limited to much lower concentrations. The specific volume remains practically unchanged. The solute is here, as shown by various methods, either present in the form of very elongated or plate-like sub-microscopic units of a crystalline character, or consists of the thread-like giant molecules of a high molecular compound. A third group shows also very high viscosities and a non-linear increase at low concentrations, but the density of the solute is apparently reduced. Other methods indicate the presence of micelles or of aggregates of primary units of the first two types, including great amounts of the solvent, and varying in shape.

There exist, of course, intermediate types and special cases as well, but all show (from groups 1 to 3 increasing) one common fundamental peculiarity: the viscosity is not a function of the concentration alone but, for the same dry weight of the solute, can change up to 100 times or more under influences such as the temperature, the solvent, additional substances, velocity of flow, history of the system, etc. As a matter of fact, above a certain limit of concentration which varies in different cases from less than 0·01 per cent to about 5 per cent, no change in a colloid solution is possible without affecting the viscosity.

Considering these facts, and in view of what was stated at the beginning about aggregation and solubility of the high polymers, one comes to assume several main mechanisms for the influence of concentration:—

1. The viscosity is increased by the presence of large individual particles which disturb the flow; the amount can be calculated hydrodynamically as additional dissipated energy: (*a*) the viscosity in dilution increases proportionally to the concentration, and the factors depend on the particle shape and orientation as discussed above; (*b*) at higher concentrations the separate ranges of disturbance begin to overlap, and this interference, as well as, (*c*), the mutual influence of the particles on their orientation, can be calculated in principle by an extension of the hydrodynamic method.

2. For still higher concentrations when collisions between the units of the solute become more frequent, the deviation from the linear relation loses the nature of a correction and becomes the main effect; new methods of calculation will have to be opened up, which should proceed kinetically or statistically as pointed out for the viscosity of pure liquids.*

This picture should represent truly the conditions for the first group, e.g. dispersed rigid spheres, or rods and plates of moderate axial ratio, but very elongated particles or threads must behave differently. Expressing the mean distance between them as the ratio of the distance to the particle dimensions, one finds low values (=low concentration) when the two dimensions perpendicular to the long axis are considered, but a very high value parallel to it. Accordingly the mechanisms 1 (*b*) and 1 (*c*) should occur at lower concentrations, while 2 will have already set in when the viscosity due to the other factors is still rather low.

3. Concentration further changes the structure of the solution as a whole and, by diminishing the activity of the solvent, reduces the solvation jacket round the particles. This first counteracts the other effects of concentration to some extent, but at the same time, as the jacket becomes more defective, the activity of the solute increases until the denuded parts join and, closing the jacket of the aggregate again, restore the stability, but now on a lower level. In accordance with the negligible solubility of the unsolvated material, aggregation may occur at any concentration and upset what was stated under 1 and 2 as based on the presence of stable single particles. Aggregation can affect (*a*) the actual concentration, by including solvent: by forming anything from rather dry crystals to a very loose network, practically any amount of liquid can be immobilized†; (*b*) the axial ratio, which might be either reduced if the particles pack side by side or if almost symmetrical flocks appear, or be increased by rods sticking end to end; (*c*) the form: these aggregates may remain as new independent units (micelles), or they may join each other to form microscopic structures.

What are the possibilities of deciding whether in a solution of high polymers separate units are present to which the observed properties can be ascribed, whether some sort of contact exists between them, or whether they are definitely united to aggregates? Determining η_{sp}, one usually finds a group of solvents yielding the same low specific viscosity if extrapolated to infinite dilution, and therefore indicating single particles. Less good solvents, as they are frequently called, show by a much higher viscosity the presence of interaction of some kind.‡ Changes of η_{sp} with the temperature sometimes give a rather confusing picture.

* It is, perhaps, not only a coincidence that, as shown below, the most successful empirical equations for high concentrations are of the logarithmic type. On the other hand, starting from the viewpoint of a mechanical equilibrium (thus related to Prandtl's idea), an interesting attempt was recently made by Goodeve which, although carried out in a very elementary way, with very special assumptions, is of great stimulating value.

† Assuming closest packing, the space included in the aggregate is 0 for parallelepipeds, 5 per cent of the particle volume for cylinders and 30 per cent for spheres.

‡ Between these two extreme cases there actually exists a viscosity minimum, belonging to the minimum solvation of the single molecules before they start to aggregate.

While marked decreases show the presence of micelles or of extended aggregates, small changes can also be found in molecular disperse systems due to the influence of the thermal agitation on the orientation or to alterations in the degree of coiling up. Independence of the temperature, on the other hand, can arise from a partial compensation of some of these effects. The same is true for the decrease of viscosity with the rate of shear. In many cases, however, doubts can be eliminated, or one or the other possibility supported by changing solvent, adding traces of other substances, or by considering the general influence of temperature. Valuable assistance can often be obtained from osmotic-pressure measurements.

Diffusion and sedimentation measurements rather give evidence of the transition from the diluted state to the mechanism referred to under 2. Measuring the diffusion constant of methyl and acetyl cellulose fractions, Polson found marked deviations from the ideal gradient curve for a molecular weight of 53 000 in 0·5 per cent solution and for 100 000 in a 0·25 per cent solution. This goes to support McBain's claim that most of the viscosity data in the literature on cellulose derivatives have been obtained with associated material, and are, therefore, not conclusive for molecular properties. A very striking effect was revealed by Signer for a polystyrene of molecular weight 270 000 which showed a reduced sedimentation velocity with as low a concentration as 0·0001 g./c.c.

Considering this scale of concentration effects, one recognizes that a general attack on this problem must be extremely difficult and that separate attempts need a very thorough analysis of the underlying cause. This can be much more easily achieved by such theoretical work as was recently carried out by Guth, Gold and Simha. They adopted for the extension of the Einstein formula a method which was introduced by von Smoluchowski to calculate the interaction of spheres in parallel translational motion. The flow round a first particle, still disturbed when the fluid reaches the second particle, must be compensated there by an additional flow in order to fulfil the condition of adhesion to this particle This additional flow reacts in turn on the first sphere and has, moreover, to disappear at infinity. The reflexion can be repeated at will, but a single one is enough to obtain the quadratic term. The energy dissipated in the modified flow of the suspension is now calculated by means of (1) and compared with the same formulated for a pure liquid, yielding the formula

$$\eta_{sp} = 2{\cdot}5c + 14{\cdot}1c^2. \qquad \ldots\ldots(9)$$

It seems worth noting that in (9), as in (2), η_{sp} is independent of the radius of the spheres, but that in this case two minor factors have been neglected, one depending on the size of the particles, the other on the forces acting between them.

The unsolved problem of the orientation prevented the same authors from calculating the interaction of large ellipsoids in a consistent way. In order to get some idea, an idealized case was considered, i.e. elongated ellipsoids oriented parallel to the direction of deformation and spinning undisturbed by each other. They obtain as a result

$$\eta_{sp} = h\left(\frac{f}{2\log 2f - 3} + 2\right)c + k\frac{f^3}{(2\log 2f - 3)^2}c^2, \qquad \ldots\ldots(10)$$

where h and k are near unity, h = constant, k depending on the rate of shear. For smaller rods in Brownian movement, Guth expected by analogy a formula of the type

$$\eta_{sp} = \alpha_1 f^2 c + \alpha_2 f^4 c^2 + \ldots, \qquad \ldots\ldots (10a),$$

where the factor $\alpha_1 f^2 = \gamma$ is, of course, the same as in the formulae (5) or (6) above. With regard to (10) and (10*a*), one recognizes a very steep increase with the concentration for greater axial ratios and consequently a shrinking of the linear part of the relation to insignificance.

Formula (9) was proved by the author and Goldschmid by means of models (glass spheres $r = 50$ to $150\,\mu$) and found strictly valid up to 6 per cent in the Couette viscosimeter. Measurements with capillary viscosimeters gave somewhat lower values than (9), an effect which could be traced to the influence of inertia. When (10) was used to calculate the factors γ from Sverak and the author's viscosity measurements with models of rod suspensions, a fair agreement with Jeffery's maximum values (4) could be obtained. Formula (10) described also the viscosity concentration curves so long as the quadratic term did not exceed the linear by more than about 50 per cent, i.e., so long as the character of a correction was preserved. For (10*a*) some proof was given by Kratky and Saito.

The influence of aggregation on (2) was recently investigated by Broughton and Windebank. They found a fair agreement in cases where the aggregate could be assumed to be either small or easily broken up by the shear, but rather marked deviations in other cases. The reason for that is not given by the authors. As one reason, however, the amount of included liquid could be suggested, while similar unpublished experiments by the author indicate further the influence of the asymmetrical shape of the aggregates. It is to be regretted that no microscopic tests were reported in the paper.

An experimental investigation based on hydrodynamics was carried out by the author, Huber and Mark. It concerns the other end of the scale of concentration effects, dealing with the mechanical hindrance (immobilization) of the solvent between the meshes of the aggregates. It was possible to show that the resistance of slowly moving perforated spheres and ellipsoids approaches the resistance of the full body, thus indicating that the liquid inside was dragged along in spite of the free communication. This fact, although for experimental reasons established for progressive motion, seems to support strongly the idea of an immobilized liquid, increasing the effective concentration.

These results are related to a rough calculation by Kuhn, who showed that a great number of small spheres spread out over the surface of a very much larger imaginary sphere may cause a greater resistance, if the large body built in such a way is penetrated by the liquid than if it were full. McBain's idea, however, to take as a measure of the immobilized volume of liquid the resistance of dropping ellipsoids in comparison to that of a sphere of the same volume, is not justified. This ratio, if at all, can only be an analogy to the ratio between Einstein's factor 2·5 and the equivalent for a suspension of ellipsoids.

Between the two attempts to calculate the hydrodynamic interaction and to ascertain the mechanical fixation of liquid lies the field of empirical equations which deal with combined concentration and aggregation effects. There are a number of well-known formulae, due, for instance, to Arrhenius, Baker, Berl, Fikentscher, Philipoff and others, which are quite useful in conjunction with the experimental facts. Starting from quite different points of view, these equations are not greatly different in their internal structure. As pointed out by Guth and Mark and later by Sakurada and by Bredee and Booys, they can all be considered as a series of powers in the concentration, cut short after one or a few members. Their coefficients, of course, are different, but within one series in many cases related one to another by a simple relation (Sakurada).

It seems difficult, however, especially having regard to the complex nature of the concentration effects, to draw any conclusions about the properties of the solute from such data in concentrated solutions. Nevertheless, stimulated by the theoretical development, many attempts have been made recently to interpret the constants in these equations. They all proceed by formulating the empirical relations in such a way as to render them identical with the theoretical formula for low concentrations and to derive from this the physical meaning of the constants. This is, of course, inconsistent, as any formula loses its meaning outside the assumptions for which it was derived. It follows, moreover, from the general considerations given above, that different stages of concentration cannot be treated in the same fashion. The introduction of new constants makes a formula more adaptable but makes it increasingly difficult to come to any possible interpretation. One of the logical conceptions is the equation of Fikentscher and Mark:

$$\eta_{sp} = k\left(\frac{\zeta}{1-\zeta}\right)$$

(k = constant, ζ = concentration of the solute + solvation), but it does not cover a great range. The most logical way to build up a formula for a great range would be that of Philippoff and Hess. When they write

$$\eta_{sp} = \left(1 + \frac{[\eta]}{8}\right)^8, \quad [\eta] = \eta_{sp}/c_{\lim.\, c=0},$$

the latter expression is experimentally determined and represents, therefore, a material constant. $[\eta]$ is, for instance, equivalent to α_1/f^3 in (10 *a*) (including hydrodynamic, solvation and other facts), and raising it to the same power as the accompanying term of c, Philippoff actually follows Guth's proposal. That the exponent 8 covers so many of the experimental facts may be rather a coincidence.

In the course of the laborious attempts to compare the various {viscosity, concentration} curves and to correlate them with a suitable formula, some interesting facts with regard to their applicability have been disclosed. It has been known, for instance, for a considerable time that the Arrhenius and Fikentscher equations:

$$\log \eta_{rel.} = k \,.\, c \quad \text{and} \quad \log \eta_{rel.} = \left(\frac{0{\cdot}75k^2}{1 + 1{\cdot}5kc} + k\right)c$$

have about the same number of satisfactory applications, but practically none in common. Houwink and, in more detail, Bredee and Booys have pointed out that it is the great majority of synthetic polymers which obey the Arrhenius equation, whilst most of the natural high polymers follow Fikentscher's or Philippoff's formula. The same authors then go a step further and show that this division is due to the fact that most of the natural high polymers can be assumed to be in solution as micelles or molecular aggregates, while the synthetic compounds more readily dissolve molecularly, and that it is actually this difference in the state of solution which is decisive for the applicability of one or the other formula. One cannot follow the authors further in the interpretation of their constants for the reasons stated above, but there is another point of interest. The comparison of the various formulae mentioned was carried out over a range of concentration from about 1 to 30 per cent. As we saw in a former section, even the lower end of this range must be considered as rather concentrated for molecules which are sufficiently elongated, and the extremely high viscosities of a number of the solutions show that the solute must already be present in the form of aggregates (micelles). Mechanism 2 will prevail for the increase of viscosity with concentration in those solutions so long as no further aggregation starts, while the amount of liquid held by the particles during the same period must decrease continuously. The {log viscosity, concentration} curve of such a system should therefore become less steep in the range considered, while that of solutions of other types, for which aggregation only becomes considerable above 1 or a few per cent, should rise more steeply. If one follows the course of viscosity in such latter cases higher up, one finds that artificial resins, for instance formaldehydes, yield a curve which is bent upwards in contrast to that of polystyrenes, which bend towards the abscissa. This means, according to the view set out in this report, that aggregation of the long-shaped styrenes occurs in a way in which reduction of the axial ratio weakens the other effects of concentration. This influence will be more marked in the case of single flexible molecules, which are able to grow more easily into parallel bundles than stiff rods (micelles) could do, which rather stick together in the random orientation in which they meet.

Bredee, Booys and Houwink's discrimination of the high-molecular solutions consequently seems justified only in view of the concentration range they chose, thus providing our postulated mechanisms for the concentration effects with interesting examples. At another concentration range their observation might have been inverse, fitting into the present stage of the long-standing discussions as to whether the high polymers exist, either in part or entirely, as micelles or as molecules in solution.* The present answer is that they are giant molecules and that it depends, further, entirely on the concentration, temperature and on the solvent where the transition zone between the single and the associated state must

* This discussion was actually a repetition of the still older question whether the colloidal state is a general possibility or a special feature of some substances, a question which has to be answered in the same way by the dependence on the external conditions.

be looked for. As a proof of this view a steep rise in viscosity is to be expected for molecules with strong aggregation tendencies at concentrations of less than 1 per cent; the few measurements existing in sufficient dilution confirm this requirement thoroughly (Guth and Mark and Fikentscher).

It seems of great importance that our views, developed for systems where the single factors could be studied and understood separately, apparently hold quite well up to concentrations where a great number of solute particles remain in contact for a considerable period of time. The superstructure built in this way changes the mechanical nature from that of an ordinary liquid to that of a non-Newtonian or of a highly elastic body.* It is very satisfactory that those systems can be approached by steps, starting from dilute solutions, and that a consistent way is thus opened to study these phenomena which are so important throughout the whole living world. To the technical chemist, moreover, this steady transition opens up the possibility of obtaining valuable information and guidance about the structure of the fibre or the plastic material which he is about to precipitate, from the shape of the molecules or the type, the size and the strength of the aggregates already formed in solution.

Lastly, it must be emphasized that in this report it has not been possible to discuss papers dealing with systems in which electrical charge plays an important part. This involves, however, no change in principle, but mere modifications in the material here discussed. While it was thus necessary for this outline to remain incomplete in many respects, it has entered in considerable detail into the field with which it deals. This seemed to be necessary having regard to the unusual complication of viscosity problems which depend on so many different factors, and the main object has been to show in a constructive way the important progress achieved during the last few years.

* Non-Newtonian behaviour in diluted systems is to be traced to orientation effects, as discussed previously.

THE DEVELOPMENT OF THE TEACHING OF EXPERIMENTAL PHYSICS IN BRITISH UNIVERSITIES

By ALLAN FERGUSON,
Queen Mary College and Cavendish Laboratory, Cambridge

THE function of university teaching is threefold—to impart existing knowledge; to lead the student to the frontiers of existing knowledge and to train him in those habits of thought and action which may result in his making some contribution, however modest, to the extension of those frontiers; and to develop character. This last-named function, by far the most important of the three, is, in our day at least, best performed indirectly. To some generations, didactic appeals have their value; but in an age when, with one unique exception, oratory is conversational, and the impressive pronouncements of Mrs. B. are apt to excite ribald amusement rather than admiration, the appeal must be indirect; example is decidedly better than precept.

In evoking qualities which are—ten years ago, one would have written universally—recognized as desirable in man the social animal,—a disinterested love of truth, an appreciation of beauty and of craftsmanship, the development of exact and logical habits of thought, the cultivation of ingenuity and resourcefulness,—a training in experimental science takes a high place. The case for laboratory teaching in experimental physics has been put with characteristic vigour and enthusiasm by Kelvin, though in these uncertain days we may possibly be inclined to qualify some of his downright statements:

"There is one thing I feel strongly in respect to investigation in chemical or physical laboratories—it leaves no room for shady, doubtful distinctions between truth, half-truth, whole falsehood. In the laboratory everything tested or tried is found either true or not true. Every result is *true*. Nothing not proved true is a *result*;—there is no such thing as doubtfulness. The search for absolute and unmistakable truth is promoted by laboratory work in a manner beyond all conception. It is a kind of work in which also patience and perseverance are promoted to a most marked degree. No labour must be shrunk from; everything must be carefully done. There is this which is satisfactory about it; that perseverance is sure to be rewarded. There is no failure in physical science. We do not always find the particular thing looked for; we often find that what we looked for does not exist, or that something else exists very different from what we expected to find; but that something is to be found in any investigation entered upon with intelligence and pursued with perseverance, is a certainty . . . . Every additional knowledge of the properties of matter is of value.

"A large part of the work of a physical or chemical laboratory must be measurement. That might seem rather trying work; 'harsh and crabbed' shall we say? . . . But . . . the difficulties to be overcome in physical science in mere measurement are teeming with interest. Properties of matter, or forces to be contended with, oblige us to be always digressing. We cannot go on saying—'We will think of nothing but the object before us'. Every person who aims at one object of course perseveres until he attains it; but he keeps his mind open until he can return to some other object never thought of at first, but which thrust itself on him as a difficulty occurring in the pursuit of the first object. The very disappointments in attaining objects sought after in the investigations of physical science are the richest sources of ultimate profit, and present satisfaction and pleasure, notwithstanding the difficulties and disappointments contended with."

This passage clearly illustrates Kelvin's notion of a *teaching* laboratory. It is a laboratory where even junior students are "put at once into investigations", and it developed on these lines quite naturally from the fact that Kelvin needed, soon after he entered the Chair of Natural Philosophy in 1846, experimental data and constants which, at that time, did not exist, and volunteers from the ordinary class came forward to assist in the search for such constants. Kelvin's methods were not those which have since developed; there was an engaging lack of system about them which was possible only in the days of relatively small laboratory classes. Ramsay has put it on record that his first exercise in the laboratory was to take the kinks out of a bundle of copper wire. Having accomplished this feat, at the expense of a week's work, he was given a quadrant electrometer and set the task of measuring various potential differences, including the potential difference between the internal and external coatings of a child's balloon, blackleaded internally and externally, and filled with hydrogen.

It must not be imagined that these views of the value of laboratory teaching and investigation were universally held at the time at which they were put forward. 'The Conflict of Studies' is not a book that is widely read today, but it had its vogue, and its distinguished author expressed views which, odd though they may seem to this age, represented the opinions of a not unimportant section of the academic world of the eighteen-seventies. Writing of that urge to see and do things for one's self which is at the very heart of experimental science, he remarks (*à propos* of the guinea-and-feather experiment), "It may be said that a boy takes more interest in the matter by seeing for himself, or by performing for himself, that is, by working the handle of the air-pump; this we admit, while we continue to doubt the educational value of the transaction . . . . It may be said that the fact makes a stronger impression on the boy through the medium of his sight, that he believes it the more confidently. I say that this ought not to be the case. If he does not believe the statements of his tutor—probably a clergyman of mature knowledge, recognised ability, and blameless character—his suspicion is irrational, and manifests a want of the power of appreciating evidence, a want fatal

to his success in that branch of science which he is supposed to be cultivating ".

Such was the simple rationalism of Isaac Todhunter, at a period when changes of the first importance in his university were taking place; probably his conservative attitude was emphasized by the alarming recommendation of a Syndicate (appointed in 1868) that a Professorship of Experimental Physics should be established. Maxwell's light-hearted warning, " If we succeed too well, and corrupt the minds of youth till they observe vibrations and deflections and become Senior Optimes instead of Wranglers, we may bring the whole university and all the parents about our ears ", was a fair expression of the attitude of many of the more conservative-minded educationists of the period.

The interesting side of this view is that, forgotten though it may be today, it was held, and widely held, within living memory. Indeed, the present structure of university teaching on the experimental side of physical science, with its schedules, its benches and tables on which the essential components of an experiment are laid out in order, and its far too infrequent demands on the ability of the student to develop what capacity he may have for simple constructional work, has all evolved within the lifetime of the founder fellow of the Physical Society, Sir Ambrose Fleming. The development of students' laboratories for the teaching of experimental chemistry was earlier, but not a great deal earlier; and physical laboratories (in which much chemical work was also carried out) tended, during the seventeenth, eighteenth and early nineteenth centuries, to become concentrated in the hands of wealthy, or well-endowed, private investigators. So it was with Boyle, when he took " a long Glass-Tube, which by a dexterous hand and the help of Lamp was in such a manner crooked at the bottom that the part turned up was almost parallel to the rest of the tube, and the Orifice of this shorter leg of the Siphon (if I may so call the whole Instrument) being Hermetically seal'd, the length of it was divided into Inches (each of which was subdivided into eight parts . . .)". Having enclosed a known volume of air in the shorter leg of the tube, and having taken care " by frequently inclining the Tube " that the air therein should be at the same pressure as that of the external air, " we began to pour Quicksilver into the longer leg of the Siphon till the Air in the shorter leg was by condensation reduced to take up but half the space it possessed (I say, *possess'd* not *fill'd*) before; . . . and . . . observed, not without delight and satisfaction, that the quicksilver in that longer part of the Tube was 29 Inches higher than the other". So it was with Cavendish; so with Priestley, whose work would have been almost impossible but for the financial help afforded him by a group of friends who saw the promise of his work. Priestley's was a restless and versatile intellect which ranged widely over the problems of his age, and it is interesting to compare his views with those expressed by Todhunter. Writing in 1790 he observes that " natural science is very little, if at all, the object of *education* in this country . . . . If we wish to lay a good foundation for a philosophical taste, and philosophical pursuits, persons should be accustomed to the sight of experiments, and processes, in *early life*. They

should, more especially, be early initiated in the theory and practice of *investigation*, by which many of the old discoveries may be made to be really *their own*; on which account they will be much more valued by them. And, in a great variety of articles, very young persons may be made so far acquainted with everything necessary to be previously known, as to engage (which they will do with peculiar alacrity) in pursuits truly original". The young person of today has a longer row to hoe than the youth of 1790, but he starts with more advantages, and Priestley's outlook is fundamentally sound.

Probably the most extensive private physical laboratory—one, too, which served as a centre of attraction for workers from many countries—was that of Gustav Magnus, a laboratory which is a link between the age of the private investigator and that of the university teaching laboratory, for the magnificent collection of instruments formed by Magnus was purchased for the University of Berlin, and became the nucleus of the teaching laboratory of the university.

Until the eighteen-sixties, or thereabouts, the student of physics must perforce find a place in the laboratory of some eminent private investigator if he desired training in experimental method. This was true equally for physicists and for chemists; indeed, then as now, there was considerable overlap in their studies, and that differentiation, which developed in the nineteenth century and which is now rapidly disappearing, had hardly appeared in the period with which we are dealing. One outstanding difference did, and does, exist; chemists are a more loquacious race. And, if we would seek a vivid picture of the conditions of training in the pre-teaching-laboratory period, it is to the chemists that we turn. No better account can be found than that given in the autobiography of Sir Robert Christison, who for many years held the Chair of Materia Medica in the University of Edinburgh. Christison's life (1797–1882) almost spanned the century, and he lived every day of his life. His lively pen chronicles a wide experience. After graduating at Edinburgh, he studied, about 1820, in London and in Paris, and wisely made the most of his opportunities. Nothing escapes him. The contemporary state of medical studies in the two great cities is portrayed in masterly fashion. Now, he sketches the French "Diligence Célérifère"; again he mystifies an inquisitive English-speaking Frenchman by "kittling his lugs wi' a wee braid Scotch". He attends the sittings of the scientific department of the Institute and gives us admirable thumbnail sketches of the lectures and the appearance of Ampère ("a shy, simple-hearted, downward-looking man") of Haüy, Gay-Lussac, Thénard, Laplace, Arago, and a host of other notabilities. His method may be illustrated by his impression of Laplace, "a slender figure, small sharp features, a high prominent wide brow, white locks hanging straight down his temples, and a benevolent droop of the lower lip. I heard him speak only once, and but a few words—with a bitterness which I was far from expecting in a man of so engaging an expression of countenance, and one belonging so eminently to the *animis celestibus*. A discussion arose on a mathematical paper by a young aspirant. At the close the President asked, 'Qu'est-ce qu'en pense

M. de Laplace?' To which the philosopher replied, 'Monsieur le Président, je n'en pense pas du tout!'".

Christison worked in the laboratory of Robiquet, the eminent pharmacist, and discoverer of some of the alkaloids of opium. Christison's description of the laboratory, of his daily work and his colleagues, his successes and mischances, is given in his accustomed lively fashion, and is one of the best accounts extant of the laboratory life of a student during this period. We see the dominating figure of Robiquet—quick, active, keen-eyed and keen-tempered. One day Christison, ignorant of the French habit of weighing with an equipoise, removed it, as being merely a relic of some previous work, so spoiling the labours of a whole forenoon. "Ah! qu'est-ce que c'est que vous avait fait? La tare est perdue—et votre procédé! mais, enfin! il faut recommencer. Voulez-vous venir demain? Non?" And the Sunday-observing young Scot, shame for his error overcoming his scruples, performed his first successful analysis on a Sunday.

The autobiography is a quarry for material for the study of the academic and social life of Christison's period that has been far too little used. It is a misfortune that Kelvin has left so slight a record of the months which he, in similar circumstances, spent in Regnault's laboratory in 1845, where his experimental work seems, for the most part, to have consisted in "working the handle of the air-pump".

But a new era was fast dawning. About 1830, Thomas Thomson had opened in Glasgow a laboratory for students of chemistry. Kelvin, soon after his own appointment in 1846, had seized upon a disused wine-cellar, in which enthusiastic volunteers attacked the work already mentioned. In the late 'sixties R. B. Clifton proposed a course in physical experiments as part of a degree training and established the Clarendon Laboratory; King's College in London and the Massachusetts Institute of Technology in the United States were the first to put into practice Clifton's suggestion; the Cavendish Laboratory was opened in 1874.

The Massachusetts Institute of Technology was the first institution to adopt the modern plan of putting out on a table the essential pieces of apparatus for an experiment, together with a brief account of the experiment. In connection with this work, E. C. Pickering, who was the head of the department of physics, published a treatise on practical physics which is a model of its kind, and, with a few deletions and additions, might well serve as a present-day elementary handbook. Physics is sometimes a Cinderella in an institute of technology, and the time to be given to laboratory work in physics is restricted. Pickering's choice of experiments was very carefully considered, and it is interesting to note that, of the two experiments which he selected to illustrate photometry, the experiment on the comparison of candle-power has remained an integral part of every elementary laboratory course, whereas the vastly more important experiment on illumination dropped out of use and has been revived only during recent years. There is a real need for a *simple* form of illumination photometer suited to the needs of students of intermediate grade, and Pickering's primitive form—

a long rectangular box blackened inside, closed at one end by a grease-spot screen, and furnished with a lamp or candle whose distance from the screen may be varied, fulfils its task as well as, if not better than, some more ambitious types.

With slight variations, the method of practical teaching in physics is, in the main, that instituted by Pickering. Is there much to criticize in the working of the system? As regards the content of the work, I hardly think there is. In the various undergraduate grades, intermediate, pass and honours, the usual choice of experiments is well suited to the needs of the average student, and may serve to stimulate the interests of the man who is above average quality. In the intermediate and pass grades the type of experiment is to a large extent standardized, though there is always room and need for a sprinkling of experiments which shall have that touch of novelty which serves to keep alive the interest both of student and demonstrator. At the intermediate grade, the teacher who is seeking this element would be well advised to consult the pages of the editions of the *Science Masters' Handbook*; they are a veritable storehouse of hints towards ingenious and simple experiments of all types. At the pass, Part I, or general honours stage there are many possibilities of instituting experiments with relatively simple methods and materials, which shall put the student in the position of an investigator. In selecting such problems, each teacher must draw upon his own experience, and I quote, with a certain amount of hesitation, one or two experiments which seem to me to possess that quality. Every student of physics must, at some time or other, have determined a surface-tension by the capillary-rise method, or by the pull on a plate whose edge just touches the surface of a liquid. Such experiments, new through they may be to any individual student, are apt, with the passage of years, to become endued with a measure of something very like senility. It gives to these experiments an entirely new zest if the student is asked to compare the results of the capillary-rise experiment with those obtained by the use of one or two micro-methods, say, by pushing a thread of liquid along a horizontal capillary until the meniscus at the free end of the capillary is plane, or by applying pressure to a thread of liquid contained in a vertical U-tube made of two capillaries of different diameters until the two menisci are at the same level.

So, too, with the balance method. Instead of equilibrating the surface-tension pull by means of weights, the equilibration may be brought about by raising the vessel containing the liquid through a known height such that the upward thrust due to buoyancy just balances the surface-tension pull. The plate usually employed may be replaced by a cylinder, solid or hollow, with its axis vertical.

The elastic properties of materials in the form of thin wires may very profitably be studied by means of a simple modification of 'sGravesande's method, in which the wire under test is stretched horizontally between two fixed points and loaded at its centre.

The Newton's rings experiment furnishes another example of an experiment

which is none the worse for rejuvenation, and a good student may usefully employ himself in obtaining and studying the ellipses produced when a converging lens of low power is supported on a glass beam bent in the usual fashion.

At the Part II or special honours stage considerable latitude of choice exists. The student should be thoroughly well drilled in handling a few selected instruments of precision—interferometers, the Kew magnetometer, precision potentiometers. He should be acquainted with standard A.C. bridge methods, with some of the fundamental experiments in radioactivity, with the properties of valves and of oscillographs. And he should be encouraged to develop any power which he may possess of achieving accurate and important results with relatively simple apparatus.

It is possible, for example, with such apparatus, and a furnace of simple design, to make accurate quantitative studies of the variation with temperature of the capillary constants of such interesting substances as liquid crystals. The surface properties of these substances are specially noteworthy, as they furnish examples of liquids whose surface tensions, over a limited range of temperature near the transition point, *increase* with increasing temperature.

Nor is it sufficient merely to obtain a value for the measure of a physical quantity. The property, once measured, must be studied in the light of Manning's remarks to the youth who had to leave him in order to catch a train which would take him to his destination an hour earlier: " And, my son, having gained your hour, what will you do with it? "

At all stages above the intermediate, some measure of workshop practice and of glass-blowing is essential. Not that this is as common as it should be. There is a tendency at the present day to develop a race of " investigators " who call helplessly for the workshop mechanic when anything goes wrong, who are incapable of making a good T-joint, much less of constructing a simple water-pump —a feat which should be well within the compass of any capable experimenter.

The life of a teaching laboratory turns on interest—the interest of the student and the interest of the demonstrator. A mildly cynical teacher of long experience was wont to remark that the first lesson a young demonstrator should learn is to curb his passion for demonstrating. There may be an element of truth in this, but the curbing of that passion is apt to lead to a lack of enthusiasm, and that attitude is faithfully reflected in the corresponding lack of interest shown by the student. Schedules are useful, are necessary, but they have their dangers. A new experiment, explained *viva voce*, almost invariably evokes interest. Written out, much of its interest evaporates, if the student is left, unaided, to plough his way through the pages of a sometimes-dull schedule. Fortunately, there is always some turn or application which may serve to give liveliness to a routine experiment.

I have never, for example, found students show much enthusiasm for the study of the focal lines of a thin converging lens, if they are merely asked to plot curves between v_1, v_2 and ϕ, or to verify the $\sec^2\phi$ relation. But their interest is at

once aroused when it is pointed out to them that an elimination of ϕ' between $\sin\phi = \mu \sin\phi'$, and the equation

$$1/v_2 + 1/u = (\mu\cos\phi' - \cos\phi)(1/r + 1/s)$$

—the equation to the secondary focal line—leads to an equation which we may write as

$$X^2 = -2K.X\cos\phi + K^2(\mu^2 - 1),$$

where X stands for $(1/v_2 + 1/u)$ and K for $1/(\mu - 1)f$. So that a plot of X^2 against $X\cos\phi$ yields a line from whose constants the refractive index of the material of the lens may be deduced.

The measurement of a surface tension from observation of the ascent of a liquid between two vertical plates slightly inclined to each other has always been regarded as a matter of some uncertainty, the main difficulty being due to the impossibility of exactly determining the position of the axes. Grunmach has endeavoured to overcome this difficulty by ruling a series of hyperbolas on one of the plates and adjusting the angle between the plates until the curve of the liquid coincided with one of the hyperbolas; but the results remained uncertain. The difficulty may be removed, and the method made quite accurate, by measuring the co-ordinates of a series of points on the curve of the liquid relative to any pair of intersecting lines, one of which is vertical and the other horizontal. The reduction of the results involves a determination of the constants in the equation

$$(x + h)(y + k) = C,$$

the graphical solution of which may be left to the reader.

Examples of this type may be multiplied almost indefinitely, but interesting and useful though it may be to discuss syllabuses, schedules, and programmes of work, the plain fact is that the quality of a laboratory is determined in overwhelming measure by the quality of the teacher. There will always be first-rate physics and vigorous and unflagging interest in a laboratory guided by the genius of a Dr. Searle.

REFERENCES

CAJORI, 1899. *History of Physics.*
CHRISTISON, 1885. *Life of Sir Robert Christison.* (Two vols.)
FERGUSON, 1928. *Sci. Progr.* January, p. 461.
FERGUSON and VOGEL, 1926. *Proc. Phys. Soc.* 38, 193.
HELMHOLTZ, 1903. *Popular Lectures on Scientific Subjects*, vol. I, p. 1.
KELVIN, 1894. *Popular Lectures and Addresses*, vol. II, pp. 489 *et seq.*
MAGIE, 1935. *Source Book in Physics*, p. 84.
MARCET, Mrs., 1858. *Conversations on Natural Philosophy.*
Maxwell, James Clerk, 1931. *A Commemoration Volume*, p. 112.
PICKERING, 1874. *Elements of Physical Manipulation.*
PRIESTLEY, 1790. *Experiments and Observations in Different Kinds of Air*, vol. I, p. xxix.
PRIESTLEY, 1806. *Memoirs of Dr. Joseph Priestley*, p. 92.
RAMSAY, 1908. *Essays, Biographical and Chemical*, p. 94.
Science Masters' Handbook, 1931. *Series I*; 1936. *Series II.*
THOMPSON, SILVANUS P. *Life of Lord Kelvin.* (Two vols.)
TODHUNTER, 1873. *The Conflict of Studies*, p. 17.
TURNER, 1927. *History of Science Teaching in England.*

VISCOSITY OF SUBSTANCES OF HIGH MOLECULAR WEIGHT IN SOLUTION

By F. EIRICH *

REFERENCES

ANDRADE, E. N. DA C., 1934. *Phil. Mag.* **17**, 497, 705.
ARRHENIUS, SV., 1887. *Z. phys. Chem.* **1**, 285.
BAKER, F., 1913. *J. Chem. Soc.* **103**, 1653.
BERNAL, J. D., 1937. *Trans. Faraday Soc.* **33**, 27 ; *Proc. Roy. Soc.* A, **163**, 320.
BOEDER, P., 1932. *Z. Phys.* **75**, 258.
DE BOER, J. H., 1936. *Trans. Faraday Soc.* **32**, 30.
BREDÉE, H. L. and DE BOOYS, J., 1937. *Kolloidzschr.* **79**, 31, 43.
BRØNSTED, J. N. 1931. *Z. phys. Chem.* (Bodenstein Festband), 257 ; 1938. *C.R. Lab Carlsberg* (Sørensen Jub.), **22**, 99.
BRØNSTED, J. N. and COLMANT, P., 1934. *Z. phys. Chem.* A. **168**, 381.
BROUGHTON, G. and WINDEBANK, C. S., 1938. *J. Industr. Engng Chem.* (Ind. Ed.), **30**, 407.
BURGERS, J. M., 1938. *Second Report on Viscosity and Plasticity* (Amsterdam), p. 134.
DOBRY, A., 1935. *Bull. Soc. chim. France*, **2**, 1882; 1936. *Ibid.* **3**, 1790; 1937. *Kolloidzschr.* **81**, 290.
EINSTEIN, A., 1906. *Ann. Phys., Lpz.*, **19**, 289.
EIRICH, F., BUNZL, M. and MARGARETHA, H., 1936. *Kolloidzschr.* **74**, 276 ; **75**, 20.
EIRICH, F. and GOLDSCHMID, O., 1937. *Kolloidzschr.* **81**, 7.
EIRICH, F. and MARK, H., 1936. *Ergebn. exakt. Naturw.* **15**, 1.
EIRICH, F., MARK, H. and HUBER, T., 1937. *Papierfabrikant*, **35**, 251.
EIRICH, F., MARK, H. and SIMHA, R., 1937. *Ber. Akad. Wiss. Wien*, **146**, 513.
EIRICH, F. and SIMHA, R., 1939. *J. Chem. Phys.* **7**, 116.
EWELL, R. H., EISENSCHITZ, R. and EYRING, H., 1937. *J. Chem. Phys.* **5**, 276.
EYRING, H., 1936. *J. Chem. Phys.* **4**, 283.
FIKENTSCHER, H., 1932. *Cellulose-Chem.* **13**, 58.
FIKENTSCHER, H. and MARK, H., 1929. *Kolloidzschr.* **49**, 135.
FRAMPTON, V. L. and NEURATH, H., 1938. *Science*, **87**, 468.
GANS, R., 1928. *Ann. Phys., Lpz.*, **86**, 628.
GARTENMEISTER, R., 1890. *Z. phys. Chem.* **6**, 524.
GLASSTONE, S., 1937. *Trans. Faraday. Soc.* **33**, 200.
GOLD, O., 1937. *Thesis* (Vienna).
GOODEVE, C. F., 1939. *Trans. Faraday Soc.* **35**, 342.
GUGGENHEIM, E. A., 1937. *Trans. Faraday Soc.* **33**, 151.
GUTH, E., 1936. *Kolloidzschr.* **74**, 147 ; **75**, 15.
GUTH, E. and MARK, H., 1933. *Erg. exakt. Naturw.* **12**, 115; 1934. *Mh. Chemie.* **65**, 93.
GUTH, E., MARK, H. and SIMHA, R., 1936. *Kolloidzschr.* **74**, 266.
HATSCHEK, E., 1928. *The viscosity of liquids* (London).
HERZOG, R. O. and KUDAR, H. C., 1933. *Z. Phys.* **80**, 217 ; **83**, 28.
HERZOG, R. O., ILLIG, R. and KUDAR, H., 1933. *Z. phys. Chem.* A, **167**, 329.
HESS, K. and GUNDERMANN, J., 1937. *Ber. dtsch. chem. Ges.* **70**, 1800.
HESS, K. and PHILIPPOFF, W., 1937. *Ber. dtsch. chem. Ges.* **70**, 639.
HESS, K. and TROGUS, C., 1929. *Z. phys. Chem.* B, **5**, 161 ; 1931. *Ibid.* **12**, 268, 1557.
HOUWINK, R. and KLAASSENS, K. H., 1937. *Kolloidzschr.* **79**, 138.
HUGGINS, M. L., 1938. *J. Phys. Chem.* **42**, 911 ; 1939. *Ibid.* **43**, 432.
JEFFERY, G. B., 1923. *Proc. Roy. Soc.* A, **102**, 163.
KATZ, J., 1930. *Cellulose-Chem.* **11**, 17.
KENDALL, J., 1937. *Trans. Faraday Soc.* **33**, 2.
KENDALL, J., DAVIDSON, A. W. and ADLER, H., 1921. *J. Amer. Chem. Soc.* **43**, 1481.

* *Reports on Progress in Physics*, vol. 7, p. 329.

RAEMER, E. O. and LANSING, W. D., 1933. *J. Amer. Chem. Soc.* **55**, 4319; 1935. *J. Phys. Chem.* **39**, 156.
KRATKY, O. and SAITO, G., 1935. *Cellulose-Chem.* **16**, 85.
KUHN, W., 1933. *Kolloidzschr.* **62**, 260 ; 1934. *Ibid.* **68**, 2.
KYROPOULOS, S., 1939. *J. Chem. Phys.* **7**, 52.
LANSING, W. D. and KRAEMER, E. O., 1935. *J. Amer. Chem. Soc.* **57**, 1369.
LAUFFER, M. A., 1938. *Science*, **87**, 469; *J. Biol. Chem.* **126**, 443.
MCBAIN, J. W., 1935. *Nature, Lond.*, **135**, 1033.
MCBAIN, J. W. and LAING MCBAIN, M. E., 1937. *J. Amer. Chem. Soc.* **59**, 342.
MCBAIN, J. W. and SCOTT, D. A., 1936. *Industr. Engng Chem.* **28**, 470.
MACLEOD, D. B., 1923. *Trans. Faraday Soc.* **19**, 6 ; 1925. *Ibid.* **21**, 151 ; 1936. *Ibid.* **32**, 872.
MEYER, K. H. and VAN DER WIJK, A., 1935. *Helv. Chim. Acta*, **18**, 1067.
OSTWALD, W., 1929. *Kolloidzschr.* **49**, 65 ; 1932. *Z. phys. Chem.* **159**, 375, 390 ; 1938. *Kolloidzschr.* **82**, 226.
OSTWALD, W. and RIEDEL, R., 1935. *Kolloidzschr.* **70**, 73.
PERRIN, F., 1936. *J. Phys. Radium*, **7**, 1.
PERUTZ, M. F., 1939. *Thesis* (Cambridge).
PETERLIN, A., 1938. *Z. Phys.* **111**, 232 ; 1939. *Kolloidzschr.* **86**, 230.
PHILIPPOFF, W. and HESS, K., 1936. *Z. phys. Chem.* B, **31**, 237.
POLSON, A., 1936. *Nature, Lond.*, **137**, 740 ; 1938. *Kolloidzschr.* **83**, 172 ; 1939. *Ibid.* **88**, 51.
PRANDTL, L., 1928. *Z. angew. Math. Mech.* **8**, 85.
ROBINSON, J. R., 1939. *Proc. Roy. Soc.* A, **170**, 519 ; *Nature, Lond.* **143**, 923.
SAKURADA, I., 1933. *Kolloidzschr.* **63**, 311; 1938. *Z. phys. Chem.* B, **38**, 407; *Kolloidzschr.* **82**, 345.
SCHULZ, G. V., 1936. *Z. phys. Chem.* A, **176**, 317 ; **177**, 453 ; B, **32**, 27.
SHEPPARD, S. E., 1930. *J. Rheology*, **1**, 471.
SIGNER, R., 1935. *Helv. Chem. Acta*, **18**, 701 ; 1936. *Trans. Faraday Soc.* **32**, 296.
SIGNER, R. and GROSS, R., 1933. *Z. phys. Chem.* A, **165**, 161 ; 1934. *Helv. Chim. Acta*, **17**, 59, 335.
SIGNER, R. and VON TAVEL, P., 1938. *Helv. Chim. Acta*, **21**, 535.
SIMHA, R., 1936. *Kolloidzschr.* **76**, 16; 1940. *J. phys. Chem.* **44**, 25.
VON SMOLUCHOWSKI, M., 1911. *Bull. intern. Acad. Sci. Cracovie*, no. 1 *a*.
STAUDINGER, H., 1932. *Die hochmolekularen organischen. Verbindungen* (Berlin) ; 1933. *Trans. Faraday Soc.* **29**, 30 ; 1938. *Kolloidzschr.* **82**, 129.
STAUDINGER, H. and HEUER, W., 1934. *Z. phys. Chem.* A, **171**, 129.
SVEDBERG, TH. and SJÖGREN, B., 1929. *J. Amer. Chem. Soc.* **51**, 3594.
SVERAK, J., 1938. *Thesis* (Vienna).
TAYLOR, G. I., 1923. *Proc. Roy. Soc.* A, **103**, 58 ; 1932. *Ibid.* **138**, 240.
VAN DER WAALS, J. D., Jr., 1918/19. *Proc. Acad. Amsterdam*, **21**, 743.
WARD, A. G., 1937. *Trans. Faraday Soc.* **33**, 88.
VAN WIJK, W. R. and SEEDER, W. A., 1937. *Physica*, **4**, 1073.